



```

Db 350 HPKHSHSEQHPGHHPHFAHHPHEHDTHQHPHGHHPGHHPHGHHPGHHPHCH 409
QY 61 DFQDYGFCDDPPHNOGHCHGHGPPGHLRRRGPGKPRPFHCKRQIGSVVRLPLRKGEV 120
Db 410 DFQDYGFCDDPPHNOGHCHGHGPPGHLRRRGPGKPRPFHCKRQIGSVVRLPLRKGEV 469
QY 121 LPLPEANFPSPPLPHHHPHPLKPNQPP 148
Db 470 LPLPEANFPSPPLPHHHPHPLKPNQPP 497

RESULT 2
A60488
histidine-rich glycoprotein - bovine (fragments)
N:Alternate names: autorosette inhibition factor
C:Species: Bos primigenius taurus (cattle)
C>Date: 19-Mar-1993 #sequence revision 23-Mar-1995 #text_change 07-Jul-1995
C:Accession: S35687; JJC2196; A60488
R:Sorensen, C.B.; Krogh-Pedersen, H.; Petersen, T.E.
FEBS Lett. 328, 285-290, 1993
A>Title: Determination of the disulphide bridge arrangement of bovine histidine-rich gly
A:Reference number: S35687; MUID:93351678; PMID:8348977
A:Accession: S35687
A:Molecule type: protein
A:Residues: 1-25;26-52;57-75;82-88;95-119;146-173;175-206;210-309;313-445 <SOR>
A:Note: 35S-Gln and 36S-Tyr were also found
R:Halkier, T.; Andersen, H.; Vestergaard, A.; Magnusson, S.
Biochem. Biophys. Res. Commun. 200, 78-82, 1994
A>Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII
A:Reference number: JJC2196; MUID:94220160; PMID:7309439
A:Accession: JJC2196
A:Molecule type: protein
A:Residues: 1-23;35-54, 'VK', 57-101, 'R', 'TVGEYTEG', 116, 'N', 118, 'R', 120-136;137-145;150-20
A:Experimental source: plasma
R:Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.
Thromb. Res. 60, 385-396, 1990
A>Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.
A:Reference number: A60488; MUID:91196010; PMID:2084959
A:Accession: A60488
A:Molecule type: protein
A:Residues: 1-6, 'X', 8-15 <VES>
C:Comment: This protein is a single-chained plasma protein which participates in transgl
C:Superfamily: histidine-rich glycoprotein; cystatin homology
C:Keywords: glycoprotein; plasma
F/2-113/Domain: cystatin homology <CY1>
F/122-207/Domain: cystatin homology (fragments) <CY2>
F/7-424, 60-71, 87-108, 165-346, 180-203, 258-288/Disulfide bonds: #status experimental
F/74, 107, 164, 266/Binding site: carbohydrate (Asn) (covalent) #status experimental
Query Match 52.1%; Score 503.5; DB 2; Length 445;
Best Local Similarity 49.2%; Pred. No. 1.3e-35;
Matches 91; Conservative 15; Mismatches 26; Indels 53; Gaps 6;
QY 1 HPKHSH-----HEQH-----PH-----CHH---PHAHHPH 23
Db 249 HPESHNFRCPPPLEHNHSDSPFOARAPLFPFPPLGLRCPPPGTKGNHRPPHDHSSD 308
QY 24 EHDTHRQHPGHHPGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 83
Db 309 EH-----HNHHPGHHPGHHPHGHHPHGHHPHGHHPHGHHPHGHHPH 352
QY 84 PPHCHLRGPGKPRPFHCKRQIGSVVRLPLRKGEVLPPEANFPSPPLPHHHPHPLKPD 143
Db 353 PPRHSHKRGPGKPRPFHCKRQIGSVVRLPLRKGEVLPPEANFPSPPLPHHHPHPLKPD 412
QY 144 NQPPF 148
Db 413 IQAFP 417

RESULT 3
KGZQHL
hypothetical protein Y39B6B.gg [imported] - Caenorhabditis elegans

```

```

histidine-rich glycoprotein precursor - Plasmodium lophurae
C:Species: Plasmodium lophurae
C>Date: 30-Sep-1987 #sequence revision 30-Sep-1987 #text_change 16-Jul-1999
C:Accession: A22692
R:Ravetch, J.V.; Feder, R.; Pavlovic, A.; Blobel, G.
Nature 312, 616-620, 1984
A>Title: Primary structure and genomic organization of the histidine-rich protein of the
A:Reference number: A22692; MUID:85061618; PMID:6095114
A:Accession: A22692
A:Molecule type: DNA
A:Residues: 1-351 <RAV>
A:Cross-references: GB:X01469; NID:g9997; PIDN:CAA25698.1; PID:g9999
C:Comment: There are two copies of 16-residue repeats, two copies of 17-residue repeats,
C:Genetics:
A:Introns: 23/3
C:Superfamily: plasmodium histidine-rich protein
C:Keywords: glycoprotein; tandem repeat
F/1-23/Domain: signal sequence #status predicted <SIG>
F/24-47/Domain: propeptide #status predicted <PRO>
F/48-351/Product: histidine-rich glycoprotein #status predicted <MAT>
F/59-74, 75-90/Region: 16-residue repeats
F/91-107, 108-123/Region: 17-residue repeats
F/124-138, 139-153/Region: 15-residue repeats
F/173-301, 312-331/Region: 10-residue repeats
F/40/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 25.0%; Score 241.5; DB 1; Length 351;
Best Local Similarity 47.7%; Pred. No. 1.8e-13;
Matches 42; Conservative 1; Mismatches 34; Indels 11; Gaps 3;
QY 2 PKKHSHSEQHPGHHPH-----PHAHHPHEH--DTHRQHPGHHPGHHPHGHHPGHHPGH 54
Db 184 PHHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHH 243
QY 55 HPHCHDFQDYGFCDDPPHNOGHCHGH 82
Db 244 HHHHHHHHD-----AHHHHHHHHDAHHH 267

RESULT 4
A54523
histidine-rich protein - Plasmodium lophurae (fragment)
C:Species: Plasmodium lophurae
C>Date: 15-Oct-1994 #sequence revision 15-Oct-1994 #text_change 20-Aug-1999
C:Accession: A54523
R:Irving, D.O.; Cross, G.A.M.; Feder, R.; Wallach, M.
Mol. Biochem. Parasitol 18, 223-234, 1986
A>Title: Structure and organization of the histidine-rich protein gene of Plasmodium loph
A:Reference number: A54523; MUID:86174893; PMID:3007981
A:Accession: A54523
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-140 <IRV>
A:Cross-references: GB:M15317; NID:g160331; PIDN:AAA29616.1; PID:g552196
C:Superfamily: plasmodium histidine-rich protein
C:Keywords: tandem repeat
Query Match 22.5%; Score 217; DB 2; Length 140;
Best Local Similarity 39.6%; Pred. No. 9.4e-12;
Matches 38; Conservative 2; Mismatches 34; Indels 22; Gaps 2;
QY 2 PKKHSHSEQHPGHHPHFAHHPHEHDTHQHPHGHHPGHHPHGHHPHCH-----H 46
Db 8 PHHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHHHPHHHHHPWH 67
QY 47 PHGHHPGHHPHCHDFQDYGFCDDPPHNOGHCHGH 82
Db 68 HHHHPWHHHHHHDAHH-----HHHHHDAHH 96

RESULT 5
T45059
hypothetical protein Y39B6B.gg [imported] - Caenorhabditis elegans

```





## RESULT 13

T23089      1.  
Hypothetical protein H13N06.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T23089  
R:Submitted, N.  
R:Submitted to the EMBL Data Library, October 1997  
A:Reference number: Z19673  
A:Accession: T23089  
A:Status: preliminary; translated from GE/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-515 <W1>  
A:Cross-references: EMBL:Z99942; PIDN:CAEL7070.1; GSPDB:GN00028; CESP:H13N06.5  
A:Experimental source: Clone H13N06  
C:Genetics:  
A:Gene: CESP:H13N06.5  
A:Map position: X  
A:Introns: 118/1; 156/2; 182/2; 306/1

```

Query Match      17.28; Score 166.5; DB 2; Length 515;
Best Local Similarity 28.5%; Pred. No. 6.le-07;
Matches 39; Conservative 6; Mismatches 29; Indels 63; Gaps 7;

          5 KHSSEQHPHGHHPHAH-----HPEHEDTHRQHPGHGHPHG 39
          |||
          92 HHQGCHGAHGCHGAHADCGCPYAKAAAEATAAHDHGAHDEHCHADHC-HAND 150
          |||||
          40 HHPFGH-----HPHGHRHPHGPHCHDFQDYGCDPPHNQGCCHGGPDPGHLRRRG 94
          |||
          151 HHGSHDEEDHHG-KAADHHGHSH--EDHG-----HSHGAE----- 185
          |||

          95 KGGRPFRCRIGSVR 111
          :||:
          186 -----SAKQVGDEYQ 195

```

## RESULT 14

T30173 zinc finger protein Pw1 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
C:Accession: T30173  
R:Relaix, F.; Weng, X.; Marazzi, G.; Yang, E.; Copeland, N.; Jenkins, N.; Spence, S.E.;  
Dev. Biol. 177, 383-396, 1996  
A>Title: Pw1, a novel zinc finger gene implicated in the myogenic and neuronal lineages.  
A:Reference number: Z20754; MUID:96400442; PMID:8806818  
A:Accession: T30173  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1378 <REL>  
A:Cross-references: EMBL:U48904; NID:gl197856; PID:gl197857; PIDN:AAC52770.1  
A:Experimental source: strain Sv129; limb bud  
C:Genetics:  
A:Map position: 7

```

Query Match      17.2%; Score 166.5; DB 2; Length 1378;
Best Local Similarity 36.0%; Pred. No. 1.5e-06;
Matches 31; Conservative 5; Mismatches 45; Indels 5; Gaps 2
QY 2 PKKHSHSQHPGHGHHPHAHHPHEH-DTHRQHPGHHPHGHHPGHGHHHPGHGHPHCH 60
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 793 PLDQEMRSEEPHDDXPHGQEPHDDMRPHGQEPHDDPHGQEPHGHGDEPHGQEPHGHGDEPHDK 842
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 DFQDYGCPDPPPHNQGHCCGHGPP 86
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 843 EPIDQEMRSEEPHSE---ESHGDEP 864
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15

RESOUR 13  
I49714

MHC H-2K/t-w5-linked ORF precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Aug-1996 #sequence\_revision 02-1  
 C:Accession: I49714  
 R:Han, S  
 Mol. Cell. Biol. 10, 138-145, 1990  
 A:Title: A putative transmembrane protein v  
 A:Reference number: I49714; MUID:90097821;  
 A:Accession: I49714  
 A:Status: preliminary; translated from GB/1  
 A:Molecule type: mRNA  
 A:Residues: 1-436 <RES>  
 A:Cross-references: GB>M32010; NIDG193738

Query Match	16.8%;	Score 162;	DB 2;	Length 436;
Best Local Similarity	42.9%;	Pred. No. 1.3e-06;		
Matches	39;	Conservative 2;	Mismatches 32;	Indels 18; Gaps 8;
QY	1	HPHKKH-HSHSEQHPHGH-HPHAHHPHEDHTRQHPGH----	HPHGH-HPHGHHPHGHHPH	53
DB	43	HGSHGSHSHEDPHHGSHGSHS----	EDPHHGSHGHTHS	97
QY	54	GH-HPHCHDPDYGPDPFPHPNQHCCHGHG	83	
DB	98	GHSHGSHSHSLMHG-----	CHGHAREHSHG	123

Search completed: September 23, 2004, 22:55:57  
Job time : 31.7189 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:23:24 ; Search time 101.044 Seconds  
(without alignments)  
413.849 Million cell updates/sec

Title: US-10-074-225A-5  
Perfect score: 966  
Sequence: 1 HPHKHSHQHPGHHPHAH.....PSFPLPHKPLKPDNQPPF 148

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
1: Geneseq1980s:\*  
2: Geneseq1990s:\*  
3: Geneseq2000s:\*  
4: Geneseq2001s:\*  
5: Geneseq2002s:\*  
6: Geneseq2003as:\*  
7: Geneseq2003bs:\*  
8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	966	100.0	148	5	Abb79807 Rabbit hi
2	966	100.0	148	5	Abb79806 Human his
3	966	100.0	525	5	Abb79804 Human his
4	949	98.2	525	8	Ade76897 Human pro
5	543	56.2	526	5	Abb79805 Rabbit hi
6	302	31.3	75	5	Abb07123 HRGP thro
7	254.5	26.3	245	4	Abg28019 Novel hum
8	251.5	26.0	330	4	Abg25331 Novel hum
9	248.5	25.7	406	4	Abg27250 Novel hum
10	242.5	25.1	233	4	Abg19767 Novel hum
11	241.5	25.0	351	2	Aar24393 Sequence
12	240	24.8	491	4	Abg09949 Novel hum
13	230	23.8	183	4	Abg25337 Novel hum
14	225.5	23.3	598	4	Abg23595 Novel hum
15	221	22.9	292	4	Aam18326 Peptide #
16	221	22.9	292	4	Abb37360 Peptide #
17	221	22.9	292	4	Aam30814 Peptide #
18	221	22.9	292	4	Abb32108 Peptide #
19	221	22.9	292	4	Abb22646 Protein #
20	221	22.9	292	4	Aam70490 Human bon
21	221	22.9	292	4	Aam58050 Human bra
22	221	22.9	292	4	Abg52171 Human liv
23	221	22.9	292	4	Aam05933 Peptide #
24	221	22.9	292	5	Abg40129 Human pep
25	219.5	22.7	227	4	Abg14399 Novel hum

26	215	22.3	124	4	Abg12879 Novel hum
27	210	21.7	82	4	Aam13580 Peptide #
28	210	21.7	82	4	Abb32509 Peptide #
29	210	21.7	82	4	Aam25978 Peptide #
30	210	21.7	82	4	Abb27363 Human pep
31	210	21.7	82	4	Abb18015 Protein #
32	210	21.7	82	4	Aam65721 Human bon
33	210	21.7	82	4	Aam53343 Human bra
34	210	21.7	82	4	Abg47362 Human liv
35	210	21.7	82	4	Aam01333 Peptide #
36	210	21.7	82	5	Abg35350 Human pep
37	209.5	21.7	324	4	Abg20509 Novel hum
38	209	21.6	304	4	Abg08412 Novel hum
39	207.5	21.5	79	4	Aau32905 Novel hum
40	204.5	21.2	372	4	Abb67014 Drosophil
41	196.5	20.3	449	4	Abb71938 Oestrogen
42	196.5	20.3	449	5	Aau74634 Oestrogen
43	193	20.0	266	4	Abg03494 Novel hum
44	191.5	19.8	156	4	Aam21094 Peptide #
45	191.5	19.8	156	4	Abb43409 Peptide #

ALIGNMENTS

RESULT 1  
ABB79807  
ID ABB79807 standard; protein; 148 AA.  
XX  
AC ABB79807;  
XX  
DT 25-NOV-2002 (first entry)  
XX  
DE Rabbit histidine proline rich glycoprotein His/Pro-rich domain.  
XX  
KW Histidine proline rich glycoprotein; HPGP; rabbit; antiangiogenic;  
KW cytosolic; antiarteriosclerotic; antiinflammatory; antidiabetic;  
KW synaecomolical; antiarthritic; antilucer; osteopathic; anticancer;  
KW ophthalmological; neotropic; neuroprotective; antiparkinsonian; chelator.  
XX  
OS Oryctolagus cuniculus.  
XX  
PN WO200264621-A2.  
XX  
XX  
PD 22-AUG-2002.  
XX  
PF 14-FEB-2002; 2002WO-US004336.  
XX  
PR 14-FEB-2001; 2001US-0268370P.  
XX  
PA (ATTE-) ATTENUON LLC.  
XX  
PI Donate F, Harris S, Plunkett ML, Mazar AP;  
DR WPI: 2002-666989/71.  
DR P-PSDB; AEN84910.  
PT New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for  
diagnosing or treating diseases associated with undesired cell migration,  
invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.  
PS Claim 1; Page 13; 82pp; English.  
XX  
CC The present sequence is the protein sequence of the histidine-proline-  
rich (H/P) domain of rabbit anti-angiogenic histidine proline rich  
glycoprotein (HPRG, see also ABB79805). Claimed anti-angiogenic  
polypeptides or peptides comprise: the H/P domain (see ABB79806) of human  
HPRG; an H/P domain of rabbit HPRG; a variant of these that is capable of  
inhibiting angiogenesis, endothelial cell proliferation or endothelial  
tube formation in vitro or in vivo; or a pentapeptide having the generic  
sequence given in ABB79808. Also claimed are: chemically synthesised or  
recombinantly produced peptide multimers; a diagnostically or  
therapeutically labeled anti-angiogenic polypeptide, peptide or peptide

CC multimer; a diagnostically useful HPRG-related composition, comprising  
 CC the diagnostically labeled polypeptide, peptide or peptide multimer and a  
 CC carrier; an antibody specific for an epitope of HPRG that is present in  
 CC the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any  
 CC of the domains in a way which inhibits the anti-angiogenic activity of  
 CC HPRG or the domain, or an antigen-binding fragment of the antibody; a  
 CC method for inhibiting cell migration, cell invasion, cell proliferation  
 CC or angiogenesis, or for inducing apoptosis; a method for treating a  
 CC subject having a disease or condition associated with undesired cell  
 CC migration, invasion, proliferation, or angiogenesis; a method of  
 CC stimulating or inhibiting angiogenesis in a subject; a method of  
 CC detecting the presence of HPRG or its cleavage product or its peptide in  
 CC a biological sample; isolated nucleic acids encoding the polypeptide,  
 CC peptide or peptide multimer; an expression vector; transformed or  
 CC transfected cells; a method of providing to a cell, tissue or organ an  
 CC angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its  
 CC pentapeptide, or the peptide multimer; an affinity ligand useful for  
 CC binding to, or isolating, an HPRG-binding molecule or cells expressing  
 CC the binding molecule, comprising the polypeptide, peptide or peptide  
 CC multimer, immobilised to a solid support or carrier; and a method of  
 CC isolating HPRG-binding molecule, or isolating or enriching cells  
 CC expressing HPRG-binding site or receptor, from a complex mixture. The  
 CC compositions and methods are useful in diagnosing or treating a disease  
 CC or condition associated with undesired cell migration, invasion, disease  
 CC proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic  
 CC retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or  
 CC fractures. HPRG is especially useful in inhibiting the growth of primary  
 CC tumours or metastases, and may also be used in treating neurodegenerative  
 CC diseases like Alzheimer's or Parkinson's disease. The antibodies are  
 CC stimulators of angiogenesis and are useful for promoting  
 CC neovascularization in pertinent disease states, and in various  
 CC immunoassays

XX Sequence 148 AA;

Query Match 100.0%; Score 966; DB 5; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4e-80;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HPKHSHSEQPHGHHPHAAHHEHDTROHPGHHPHGHHPHGHHPHGHHPHCH 60  
 DB 1 HPKHSHSEQPHGHHPHAAHHEHDTROHPGHHPHGHHPHGHHPHGHHPHCH 60  
 QY 61 DFQDYGCDPPPHNQGHCHGCGPPGHLRRRGGKGRPFHCRQIGSVYRLPLRKGEV 120  
 DB 61 DFQDYGCDPPPHNQGHCHGCGPPGHLRRRGGKGRPFHCRQIGSVYRLPLRKGEV 120  
 QY 121 LPLPEANFPSPFLPHKHPLKPNQPP 148  
 DB 121 LPLPEANFPSPFLPHKHPLKPNQPP 148

RESULT 2  
 ABB79806  
 ID ABB79806 standard; protein; 148 AA.

XX ABB79806;

XX 25-NOV-2002 (first entry)

XX Human histidine proline rich glycoprotein His/Pro-rich domain.

XX Histidine proline rich glycoprotein; HPRG; human; antiangiogenic;  
 XX cytosolic; antiarteriosclerotic; antiinflammatory; antidiabetic;  
 XX synaological; antiarthritic; antitumor; osteopathic; antitumor;  
 XX ophthalmological; nontropic; neuroprotective; antiparkinsonian.

XX Homo sapiens.

XX W0200264621-A2.

XX 22-AUG-2002.

PF 14-FEB-2002; 2002WO-US004336.  
 XX  
 PR 14-FEB-2001; 2001US-0268370P.  
 XX  
 PA (ATTE-) ATTENUON LLC.  
 XX  
 PI Donate F, Harris S, Plunkett ML, Mazar AP;  
 XX  
 XX WPI; 2002-665989/71.  
 DR P-PSDB; AEN84910.  
 XX  
 XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for  
 PT diagnosing or treating diseases associated with undesired cell migration,  
 PT invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.  
 XX  
 PS Claim 1; Page 13; 82pp; English.

XX The present sequence is the protein sequence of the histidine-proline-  
 CC rich (H/P) domain of human anti-angiogenic histidine proline rich  
 CC glycoprotein (HPRG, see also ABB79804). Claimed anti-angiogenic  
 CC polypeptides or peptides comprise: the H/P domain of human HPRG; an H/P  
 CC domain of rabbit HPRG (see ABB12345); a variant of these that is capable  
 CC of inhibiting angiogenesis, endothelial cell proliferation or endothelial  
 CC tube formation in vitro or in vivo; or a pentapeptide having the generic  
 CC sequence given in ABB79808. Also claimed are: chemically synthesized or  
 CC recombinantly produced peptide multimers; a diagnostically or  
 CC therapeutically labeled anti-angiogenic polypeptide, peptide or peptide  
 CC multimer; a diagnostically useful HPRG-related composition, comprising  
 CC the diagnostically labeled polypeptide, peptide or peptide multimer and a  
 CC carrier; an antibody specific for an epitope of HPRG that is present in  
 CC the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any  
 CC of the domains in a way which inhibits the anti-angiogenic activity of  
 CC HPRG or the domain, or an antigen-binding fragment of the antibody; a  
 CC method for inhibiting cell migration, cell invasion, cell proliferation  
 CC or angiogenesis, or for inducing apoptosis; a method for treating a  
 CC subject having a disease or condition associated with undesired cell  
 CC migration, invasion, proliferation, or angiogenesis; a method of  
 CC stimulating or inhibiting angiogenesis in a subject; a method of  
 CC detecting the presence of HPRG or its cleavage product or its peptide in  
 CC a biological sample; isolated nucleic acids encoding the polypeptide,  
 CC peptide or peptide multimer; an expression vector; transformed or  
 CC transfected cells; a method of providing to a cell, tissue or organ an  
 CC angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its  
 CC pentapeptide, or the peptide multimer; an affinity ligand useful for  
 CC binding to, or isolating, an HPRG-binding molecule or cells expressing  
 CC the binding molecule, comprising the polypeptide, peptide or peptide  
 CC multimer, immobilised to a solid support or carrier; and a method of  
 CC isolating HPRG-binding molecule, or isolating or enriching cells  
 CC expressing HPRG-binding site or receptor, from a complex mixture. The  
 CC compositions and methods are useful in diagnosing or treating a disease  
 CC or condition associated with undesired cell migration, invasion,  
 CC proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic  
 CC retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or  
 CC fractures. HPRG is especially useful in inhibiting the growth of primary  
 CC tumours or metastases, and may also be used in treating neurodegenerative  
 CC diseases like Alzheimer's or Parkinson's disease. The antibodies are  
 CC stimulators of angiogenesis and are useful for promoting  
 CC neovascularization in pertinent disease states, and in various  
 CC immunoassays

XX Sequence 148 AA;

Query Match 100.0%; Score 966; DB 5; Length 148;  
 Best Local Similarity 100.0%; Pred. No. 4e-80;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HPKHSHSEQPHGHHPHAAHHEHDTROHPGHHPHGHHPHGHHPHGHHPHCH 60  
 DB 1 HPKHSHSEQPHGHHPHAAHHEHDTROHPGHHPHGHHPHGHHPHGHHPHCH 60  
 QY 61 DFQDYGCDPPPHNQGHCHGCGPPGHLRRRGGKGRPFHCRQIGSVYRLPLRKGEV 120  
 DB 61 DFQDYGCDPPPHNQGHCHGCGPPGHLRRRGGKGRPFHCRQIGSVYRLPLRKGEV 120





14-FEB-2002: 2002WO-US004336.

43 HGHHPHGHHPHGHHPHCHDFQDYGFCDPPPHNQ---HCCHGHGPPPGHLRRRGGPGKGP 98

The present sequence is the protein sequence of rabbit histidine proline rich glycoprotein (HPRG), a proteinaceous chelator that can be used to inhibit angiogenesis and treat cancer. Claimed anti-angiogenic polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation *in vitro* or *in vivo*; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting or cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding molecule, or isolating or enriching cells expressing HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumors or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays.

Sequence 526 AA;

**Qy** 99 RPFHCRIGSVYRLPPLRXGEVLPLPEANFSPFLPHKHPLKPDNQPP 148  
     | | | : | | | : | | | | | | | | | | : | | | : | | |  
**D6** 449 FPHWRIGSVYOLPPLOXGEVLPLPEANFPOLLRNHTHPKPEIOPTP 498

RESULT 6  
ABB07123  
ID ABB07123 standard; protein; 75 AA.

AC ABB07123;

DT 13-MAR-2002 (first entry)

DE HRGP thrombospondin (TSP)-binding domain fragment (residues 443-517).

Thrombospondin; TSP-1; histidine-rich glycoprotein; HRGP; cytostatic;  
cardiovascular; protein therapy; gene therapy; antisense inhibition;  
ribozyme inhibition; antibody inhibition; angiogenesis; human; cancer;  
TSP.

XX Homo sapiens.

YY  
PN  
IIS2001041670-A1

XX  
PD  
15-NOV-2007

XX DE 05-DEC-2000

☐ 1  
☐ 2  
☐ 3  
☐ 4  
☐ 5  
☐ 6  
☐ 7  
☐ 8  
☐ 9  
☐ 10  
☐ 11  
☐ 12  
☐ 13  
☐ 14  
☐ 15  
☐ 16  
☐ 17  
☐ 18  
☐ 19  
☐ 20  
☐ 21  
☐ 22  
☐ 23  
☐ 24  
☐ 25  
☐ 26  
☐ 27  
☐ 28  
☐ 29  
☐ 30  
☐ 31  
☐ 32  
☐ 33  
☐ 34  
☐ 35  
☐ 36  
☐ 37  
☐ 38  
☐ 39  
☐ 40  
☐ 41  
☐ 42  
☐ 43  
☐ 44  
☐ 45  
☐ 46  
☐ 47  
☐ 48  
☐ 49  
☐ 50  
☐ 51  
☐ 52  
☐ 53  
☐ 54  
☐ 55  
☐ 56  
☐ 57  
☐ 58  
☐ 59  
☐ 60  
☐ 61  
☐ 62  
☐ 63  
☐ 64  
☐ 65  
☐ 66  
☐ 67  
☐ 68  
☐ 69  
☐ 70  
☐ 71  
☐ 72  
☐ 73  
☐ 74  
☐ 75  
☐ 76  
☐ 77  
☐ 78  
☐ 79  
☐ 80  
☐ 81  
☐ 82  
☐ 83  
☐ 84  
☐ 85  
☐ 86  
☐ 87  
☐ 88  
☐ 89  
☐ 90  
☐ 91  
☐ 92  
☐ 93  
☐ 94  
☐ 95  
☐ 96  
☐ 97  
☐ 98  
☐ 99  
☐ 100  
☐ 101  
☐ 102  
☐ 103  
☐ 104  
☐ 105  
☐ 106  
☐ 107  
☐ 108  
☐ 109  
☐ 110  
☐ 111  
☐ 112  
☐ 113  
☐ 114  
☐ 115  
☐ 116  
☐ 117  
☐ 118  
☐ 119  
☐ 120  
☐ 121  
☐ 122  
☐ 123  
☐ 124  
☐ 125  
☐ 126  
☐ 127  
☐ 128  
☐ 129  
☐ 130  
☐ 131  
☐ 132  
☐ 133  
☐ 134  
☐ 135  
☐ 136  
☐ 137  
☐ 138  
☐ 139  
☐ 140  
☐ 141  
☐ 142  
☐ 143  
☐ 144  
☐ 145  
☐ 146  
☐ 147  
☐ 148  
☐ 149  
☐ 150  
☐ 151  
☐ 152  
☐ 153  
☐ 154  
☐ 155  
☐ 156  
☐ 157  
☐ 158  
☐ 159  
☐ 160  
☐ 161  
☐ 162  
☐ 163  
☐ 164  
☐ 165  
☐ 166  
☐ 167  
☐ 168  
☐ 169  
☐ 170  
☐ 171  
☐ 172  
☐ 173  
☐ 174  
☐ 175  
☐ 176  
☐ 177  
☐ 178  
☐ 179  
☐ 180  
☐ 181  
☐ 182  
☐ 183  
☐ 184  
☐ 185  
☐ 186  
☐ 187  
☐ 188  
☐ 189  
☐ 190  
☐ 191  
☐ 192  
☐ 193  
☐ 194  
☐ 195  
☐ 196  
☐ 197  
☐ 198  
☐ 199  
☐ 200  
☐ 201  
☐ 202  
☐ 203  
☐ 204  
☐ 205  
☐ 206  
☐ 207  
☐ 208  
☐ 209  
☐ 210  
☐ 211  
☐ 212  
☐ 213  
☐ 214  
☐ 215  
☐ 216  
☐ 217  
☐ 218  
☐ 219  
☐ 220  
☐ 221  
☐ 222  
☐ 223  
☐ 224  
☐ 225  
☐ 226  
☐ 227  
☐ 228  
☐ 229  
☐ 230  
☐ 231  
☐ 232  
☐ 233  
☐ 234  
☐ 235  
☐ 236  
☐ 237  
☐ 238  
☐ 239  
☐ 240  
☐ 241  
☐ 242  
☐ 243  
☐ 244  
☐ 245  
☐ 246  
☐ 247  
☐ 248  
☐ 249  
☐ 250  
☐ 251  
☐ 252  
☐ 253  
☐ 254  
☐ 255  
☐ 256  
☐ 257  
☐ 258  
☐ 259  
☐ 260  
☐ 261  
☐ 262  
☐ 263  
☐ 264  
☐ 265  
☐ 266  
☐ 267  
☐ 268  
☐ 269  
☐ 270  
☐ 271  
☐ 272  
☐ 273  
☐ 274  
☐ 275  
☐ 276  
☐ 277  
☐ 278  
☐ 279  
☐ 280  
☐ 281  
☐ 282  
☐ 283  
☐ 284  
☐ 285  
☐ 286  
☐ 287  
☐ 288  
☐ 289  
☐ 290  
☐ 291  
☐ 292  
☐ 293  
☐ 294  
☐ 295  
☐ 296  
☐ 297  
☐ 298  
☐ 299  
☐ 300  
☐ 301  
☐ 302  
☐ 303  
☐ 304  
☐ 305  
☐ 306  
☐ 307  
☐ 308  
☐ 309  
☐ 310  
☐ 311  
☐ 312  
☐ 313  
☐ 314  
☐ 315  
☐ 316  
☐ 317  
☐ 318  
☐ 319  
☐ 320  
☐ 321  
☐ 322  
☐ 323  
☐ 324  
☐ 325  
☐ 326  
☐ 327  
☐ 328  
☐ 329  
☐ 330  
☐ 331  
☐ 332  
☐ 333  
☐ 334  
☐ 335  
☐ 336  
☐ 337  
☐ 338  
☐ 339  
☐ 340  
☐ 341  
☐ 342  
☐ 343  
☐ 344  
☐ 345  
☐ 346  
☐ 347  
☐ 348  
☐ 349  
☐ 350

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

PA (SILV/) SILVERSTEIN R L.

PI Simantov R, Silverstein RL;  
xx

DR WPI; 2002-017220/02.

PT Thrombospondin-binding molecules and fragments comprising regions of the  
PT TSP-1 binding domains of Histidine-Rich Glycoproteins, useful for  
PT treating cancers and cardiovascular diseases.  
PT

PS Example; Fig 1; 22pp; English.

The invention provides thrombospondin (TSP)-binding molecules and fragments comprising regions of the TSP-1 binding domains of His19idine-Rich Glycoproteins (HRGP). The proteins are used for the regulation of cellular growth and proliferation, such as in the accretion of new blood vessels (angiogenesis), particularly for treatment of cardiovascular disease. They may be used for the reduction of angiogenesis, including growth suppression and arrest, and apoptosis in normal development, for example in embryogenesis, and in a wide range of disorders and diseases, including those involving tumours, malignancies, neoplastic and other pathological conditions and homeostatic imbalances in the control of growth and development, in particular anal cancer, bladder cancer, small cell lung cancer, non-small cell lung cancer, bone cancer, brain cancer, breast cancer, cervical cancer, chondrosarcoma, clear cell adenocarcinoma (DES), colorectal cancer, endometrial cancer, oesophageal cancer, cancers of the eye, eyelid, Kaposi's sarcoma, kidney cancer, cancer of the larynx, leiomyosarcoma, leukemia, liver cancer, lung cancer, lymphoma, melanoma, mesothelioma, oral cancer, ovarian cancer, pancreatic cancer, prostate cancer, skin cancer, squamous cell cancer, stomach cancer, testicular cancer, thyroid cancer, hepatoma, neuroendocrine cancer, liposarcoma, head and neck cancer and a cholangiocarcinoma. The present sequence represents the TSP-binding motif of HRGP.

Sequence 75 AA;

Query Match 31.3%; Score 302; DB 5; Length 75;  
Best Local Similarity 100.0%; Pred. No. 4.1e-20;  
Matches 52; Conservative 0; Mismatches 0; Indels

QY 97 GPRPFHCRQIGSVYRLPPLRKGEVLPPEANFSPFPLPHHKHPLKPDNQPF 148

Db

1 GPRPFHCRQIGSVYRLPPLRKGEVLPLPEANFPSPFLPHHKHPLKPDNQPF 52

## RESULT 7

ABG28019  
ID ABG28019 standard; protein: 245 AA.

AC · ABG28019;

DT 18-FEB-2002 (first entry)

Novel human diagnostic protein #28010.

AA Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

AA  
PN  
WO200175067-A2.

11-OCT-2001.

XX  
PF 30-MAR-2001: 2001WO-IIS008631XX  
PR 31-MAR-2000. 2000US-00540217

PR 23-AUG-2000; 2000US-00649167:  
XX

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

XX

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

PS Claim 20: SEO ID NO 58378: 103pp: English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://ftp.wipo.int/pub/published/pct_sequences)

Sequence 245 AA:

Query Match 26.3%; Score 254.5; DB 4; Length 245;  
Best Local Similarity 48.2%; Pred. No. 2.8e-15;  
Matches 41; Conservative 2; Mismatches 39; Indels 3

D**b**      25 H YH HHHHHHHH YHR HHHH RPH RH HHYHH TTHHHHHHHHH YHQHH RH HHHHHHHHHLHH 84



QY 1 HPHKHSQHPKCHNPHANHHENHDTHQPHGHHPHGHHPHGHHPHGHHPHGH 60  
 Db 206 HHH 265  
 QY 61 DFODYGCDFPPHNOGCHGCHGPPGCHLRRGPGKGRPFYCRQIGSVYRPLPKGEV 120  
 Db 266 HH-----HHHHHCHHHHHCHHHHR-----HCH----- 290  
 QY 121 LPLPEANFSPFLPHKH 138  
 Db 291 -----HHRH 294  
 RESULT 10  
 ABG19767  
 ID ABG19767 standard; protein; 233 AA.  
 XX  
 AC ABG19767;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #19758.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS83954.  
 XX  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 XX  
 PS Claim 20; SEQ ID NO 50126; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 233 AA;  
 SQ  
 Query Match 25.1%; Score 242.5; DB 4; Length 233;  
 Best Local Similarity 35.0%; Pred. No. 3.2e-14;  
 Matches 57; Conservative 8; Mismatches 71; Indels 27; Gaps 7;  
 QY 1 HPHKHSQHPKCHNPHANHHENHDTHQPHGHHPHGHHPHGHHPHGHHPHGH 49  
 Db 52 HPHHH 111  
 QY 50 HPHGHHPHCHDFODYGCDFPPHNOGCHGCHGPPGCHLRRGPGKGRPFYCRQIGSV 109  
 Db 112 HYSHHHHHHHHXRDHDAHHHHHCHHCH-----HPRHGH-----HYVHHHGTV 160  
 QY 110 YRLPPLRKGEV-----LPLPEANFSPFLPHKHPLKPDNQPP 148  
 Db 161 -AFGALEDSAVARTGLPLPEGLCGAPTAAASAAVPPPAVPAP 202  
 RESULT 11  
 AAR24393  
 ID AAR24393 standard; protein; 351 AA.  
 XX  
 AC AAR24393;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 22-NOV-1992 (first entry)  
 XX  
 DE Sequence of Histidine-rich protein (HisRP).  
 XX  
 KW Malaria vaccine; Histidine-rich protein; cytoadherence.  
 XX  
 OS Plasmodium lophurae.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT Peptide /label= signal  
 FT Peptide 24..47  
 FT Peptide /label= pro-peptide  
 FT Modified-site 40..42  
 FT /label= potential glycosylation site  
 XX  
 PN US5116965-A.  
 XX  
 XX 26-MAY-1992.  
 XX  
 PF 26-AUG-1986; 86US-00900401.  
 XX  
 PR 26-AUG-1986; 86US-00900401.  
 XX  
 PA (SLOK ) SLOAN KETTERING INST CANCER.  
 XX  
 PI Ravetch JV, Pologe L;  
 XX  
 DR WPI: 1992-199590/24.  
 DR N-PSDB; AAR24393.  
 XX  
 PT Histidine-rich protein associated with plasmodium knob phenotype, and DNA  
 PT encoding it - useful for in vitro diagnosis of P. falciparum infection.  
 XX  
 PS Disclosure; Fig 7A-B; 29pp; English.  
 XX  
 CC Two variants of HisRP are produced by P. falciparum. One is associated  
 CC with what is referred to as "knobby phenotype" (K30) and "knobless  
 CC phenotype" (K-). The "knobby" and "knobless" phenotypes have been  
 CC implicated in cytoadherence, which is characteristic of erythrocyte  
 CC infection. It has now been found that cDNA expressing both K+ and K-  
 CC HisRP can be obtained by the use of P. lophurae HisRP expressing DNA. The  
 CC genomic clones (AAQ25532) is encoded in two exons, separating the signal  
 CC peptide-encoding sequence from the pro-sequence, confirming that  
 CC synthesis of the protein occurs via the preproprotein. Oligo. probes  
 CC synthesised to the signal peptide-encoding exon reveal multiple

CC homologous DNA sequences in the P. lophurae genome. The sequence of  
CC mature proteins is arranged in numerous tandem repeats with up to nine  
CC histidine residues in a row, similar to other Plasmodium proteins for  
CC which sequence data have so far been reported. (Updated on 25-MAR-2003 to  
CC correct PD field.)  
XX  
SQ Sequence 351 AA;

Query Match 25.0%; Score 241.5; DB 2; Length 351;  
Best Local Similarity 47.7%; Pred. No. 6e-14;  
Matches 42; Conservative 1; Mismatches 34; Indels 11; Gaps 3;  
QY 2 PKHSHSEQHPGHH-----PAAHHPHEH--DTHROHPGHHHPGHHHPGHHHPHG 54  
Db 184 PHHHHHHHHAPHHHHHHHAPHHHHHHHAPHHHHHHHCHHHHHHHHHHHHHHHH 243  
QY 55 HHPCHDFDYGCDPPPPNQGCCGHG 82  
Db 244 HHHHHHHHD-----AHHHHHHHDAHHH 267

RESULT 12  
ABG09949  
ID ABG09949 standard; protein; 491 AA.

XX AC ABG09949;  
XX DT 13-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #9940.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.  
XX PN WC200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS74136.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.

XX Claim 20; SEQ ID NO 40308; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used  
XX in diagnostics as expressed sequence tags for identifying expressed  
XX genes. (I) is useful in gene therapy techniques to restore normal  
XX activity of (II) or to treat disease states involving (II). (II) is  
XX useful for generating antibodies against it, detecting or quantitating a  
XX polypeptide in tissue, as molecular weight markers and as a food  
XX supplement. (II) and its binding partners are useful in medical imaging  
XX of sites expressing (II). (I) and (II) are useful for treating disorders  
XX involving aberrant protein expression or biological activity. The  
XX polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 491 AA;  
Query Match 24.8%; Score 240; DB 4; Length 491;  
Best Local Similarity 30.3%; Pred. No. 1.2e-13;  
Matches 59; Conservative 10; Mismatches 54; Indels 72; Gaps 11;  
QY 3 HKHSHSEQHP- HGHHPAHPHE-----HDTTHROHPGHHHPH-----GHHPGHHHPHG 49  
Db 173 HGHHPHPPPAAGHHQHGHGHSSPPAAGHHQHGHHPGHHQHRSPPHSTAHHQHGHHS 232  
QY 50 -----HHPGHHPCHDFQ-DYCPDPPP-----HNOGHCHGHG-PP 85  
Db 233 SPPAAGHHQHGHGHSSPPYQGHGHHPAAGHHQHRSHTAHHQ-HRSHQHGHQ 291  
QY 86 PGLRRRGPGKGP-----RPF-----HCRIGSVYRPLPKKGVLP 123  
Db 292 PGLHQRSPACATTTGTGTGTGHHQHPRTSTGHHHTGHHQHRSQHRSPPAQ----- 344  
QY 124 PEANFPSPPLPHKH 138  
Db 345 -----ATTSTAHHQH 354

RESULT 13  
ABG25337  
ID ABG25337 standard; protein; 183 AA.

XX AC ABG25337;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #25328.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WC200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS89524.

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.

XX Claim 20; SEQ ID NO 55696; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
XX sequences. (I) is useful as hybridisation probes, polymerase chain  
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
XX and in recombinant production of (II). The polynucleotides are also used







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:27:50 ; Search time 17.2369 Seconds  
(without alignments)  
447.085 Million cell updates/sec

Title: US-10-074-225A-5  
Perfect score: 966  
Sequence: 1 HPHKHSHEQHPHGHFAH.....PSFPLPHKHPKPDNQPPF 148

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	966	100.0	525	1 HRG HUMAN	P04196 homo sapien
2	543	56.2	526	1 HRG RABIT	Q28640 oryctolagus
3	499	50.6	396	1 HRG BOVIN	P33433 bos taurus
4	241.5	25.0	351	1 HRPX PLALO	P04929 plasmodium
5	196.5	20.3	449	1 CSUP DROME	Q9v3a4 drosophila
6	171.5	17.8	657	1 KNOB PLAFN	P06719 plasmodium
7	166.5	17.2	515	1 KE4L CAEEL	Q9xtc7 caenorhabdi
8	162	16.8	476	1 KE4 MOUSE	Q31125 mus musculu
9	160	16.6	549	1 DSX DROME	P23023 drosophila
10	155.5	16.1	774	1 AMY2 SCHPO	Q23218 schizosacch
11	154	15.9	352	1 KE4 BRARE	Q9pub8 brachydanio
12	154	15.9	428	1 PXB2 MOUSE	Q64733 mus musculu
13	149.5	15.5	659	1 HMN1 DROME	P22807 drosophila
14	149.5	15.5	815	1 PYGO DROME	Q9v9w8 drosophila
15	149.5	15.5	1002	1 CBPY SCHPO	O13849 schizosacch
16	147.5	15.3	469	1 KE4 HUMAN	Q92504 homo sapien
17	144.5	15.0	59	1 HPN HELPY	Q48251 helicobacte
18	143.5	14.9	852	1 SRCH RABIT	P16230 oryctolagus
19	142.5	14.8	473	1 KNOB PLAFN	P13817 plasmodium
20	142.5	14.8	634	1 KNOB PLAFG	P09346 plasmodium
21	142	14.7	2167	1 SHK1 RAT	Q9v4w8 rattus norv
22	141	14.6	402	1 SELP BOVIN	P49907 bos taurus
23	140	14.5	332	1 HRPI PLAFN	P05227 plasmodium
24	138.5	14.3	2161	1 SHK1 HUMAN	Q9v566 homo sapien
25	137	14.2	367	1 SEPA BRARE	Q98av1 brachydanio
26	137	14.2	419	1 GSC DROME	P54366 drosophila
27	137	14.2	620	1 EXTN TOBAC	P13983 nicotiana t
28	136.5	14.1	302	1 HYPB BRAJA	Q45257 bradyrhizob
29	136.5	14.1	661	1 KNG MOUSE	O08677 mus musculu
30	136.5	14.1	1319	1 MN1 HUMAN	Q10571 homo sapien
31	133.5	13.8	335	1 HYPE RHOCOA	P26410 rhodobacter
32	133.5	13.8	469	1 FXGA HUMAN	P55316 homo sapien
33	133	13.8	355	1 OTX1_MOUSE	P80205 mus musculu

ALIGNMENTS

RESULT 1

ID	HRG_HUMAN	STANDARD	PRT	525 AA.
AC	P04196;			
DT	20-MAR-1987 (Rel. 04, Created)			
DT	20-MAR-1987 (Rel. 04, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG).			
DE	HRG.			
GN	HRG.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Liver;			
RX	MEDLINE=86216149; PubMed=3011081;			
RA	Koide T., Foster D.C., Yoshitake S., Davie E.W.;			
RT	"Amino acid sequence of human histidine-rich glycoprotein derived from the nucleotide sequence of its cDNA.";			
RL	Biochemistry 25:2220-2225(1986).			
EN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Wakabayashi S., Takahashi K., Tokunaga F., Koide T.;			
RT	Submitted (JUL-1997) to the EMBL/GenBank/DBSJ databases.			
EN	[3]			
RP	SEQUENCE OF 214-247 FROM N.A.			
RX	MEDLINE=94245171; PubMed=8188234;			
RA	Hennis B.C., Frants R.R., Bakker E., Vossen R.H., van der Poort E.W.,			
RA	Blonden L.A., Cox S., Khan P.M., Spurr N.K., Kluit C.;			
RT	"Evidence for the absence of intron H of the histidine-rich glycoprotein (HRG) gene: genetic mapping and in situ localization of HRG to chromosome 3q28-q29.";			
RL	Genomics 19:195-197(1994).			
EN	[4]			
RP	SEQUENCE OF 19-27.			
RC	TISSUE=Plasma;			
RX	MEDLINE=P3095937; PubMed=1459097;			
RA	Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,			
RA	Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,			
RA	Hochstrasser D.F.;			
RT	"Plasma protein map: an update by microsequencing.";			
RL	Electrophoresis 13:707-714(1992).			
CC	!- FUNCTION: The physiological function is not yet known. It binds heme, dyes and divalent metal ions. It can inhibit rosette formation and is known to interact with heparin, thrombospondin, and the lysine-binding site of plasminogen. On the basis of its homology with HMW kininogen, the His-rich region of this protein may mediate the contact activation phase of intrinsic blood coagulation cascade.			
CC	!- SUBCELLULAR LOCATION: Secreted.			
CC	!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.			
CC	!- DOMAIN: In addition to having a high His and Pro content, this protein has many internal repeats. 12 tandem repetitions of a 5-residue sequence (GHHPH, consensus) form a histidine-rich region.			

Q63410 rattus norv  
P08297 glycine max  
P44526 haemophilus  
Q05738 mus musculu  
Q64213 mus musculu  
Q9f176 arabidopsis  
Q24256 drosophila  
Q9upw0 homo sapien  
Q15637 homo sapien  
Q62563 mus spratus  
P16399 pisum sativ







FT	TRANSMEM	214	234	POTENTIAL.
FT	TRANSMEM	247	267	POTENTIAL.
FT	TRANSMEM	297	317	POTENTIAL.
FT	TRANSMEM	386	406	POTENTIAL.
FT	TRANSMEM	429	449	POTENTIAL.
FT	TRANSMEM	463	483	POTENTIAL.
FT	DOMAIN	92	182	HIS-RICH.
FT	CARBOHYD	7	7	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	237	237	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	379	379	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	488	488	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	515 AA;	55500 MW; 17/D/E854F4E1DAAF CRC64;	

Query Match 17.2%; Score 166.5; DB 1; Length 515;  
Best Local Similarity 28.5%; Pred. No. 1.5e-06;  
Matches 39; Conservative 6; Mismatches 29; Indels 63; Gaps 7;

Qy	5	HHSHEQHGHGHPHAI-----	-----HPHEHDTHRQHPHGHHPHG 39
Dd	92	HHQGCHGARGGHAHADADGCGPYAKAAAEATAAHDEGHAHDHGHGAHDHG--HAHD 150	
Qy	40	HHPHGH-----HPHGHHGHGHHCHDFQYPCPPPHNQHCCHGCHGPPPPCHLRRRG 94	
Dd	151	HHGSHDEDEDHHG-HAHDHGHSH--EDHG-----HSHGAE----- 185	
Qy	95	GKGPRPFHCRIQGSVYR 111	
Dd	186	-----SAKQGVDEYQ 195	

```

RESULT 8
KE4 MOUSE STANDARD; PRT; 476 AA.
ID KE4 MOUSE
AC Q31125; Q9Z1W1;
DT 01-NOV-1997 (Rel. 35, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DD 15-MAR-2004 (Rel. 43, Last annotation update)
DE Zinc transporter SLC39A7 (Solute carrier family 39 member 1)
DE (Histidine-rich membrane protein Ke4).
DE SLC39A7 OR HKEA4 OR H2-KE4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90097821; PubMed=2294398;
RA St Jacques B., Han T.-H., Macmurray A., Shin H.-S.;
RT "A putative transmembrane protein with histidine-rich charge clusters
RT encoded in the H-2K/tw5 region of mice.";
RL Mol. Cell. Biol. 10:138-145(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: Expressed abundantly in embryonic carcinoma
CC cells, but weakly in adult tissues.
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family.
CC KE4/Catcup subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----

```

DR EMBL; M32010; AAA37767.1; -;  
 DR EMBL; AF100956; AAC69903.1; -;  
 DR MGI; 95909; H2-Ke4.  
 DR InterPro; IPR003689; Zn\_transpt\_Zip.  
 DR Pfam; PF02535; Zip; 1.  
 KW Transport; Transmembrane; Glycoprotein.  
 DR TRANSMEM 7 27 POTENTIAL.  
 FT TRANSMEM 146 166 POTENTIAL.  
 FT TRANSMEM 177 197 POTENTIAL.  
 FT TRANSMEM 222 242 POTENTIAL.  
 FT TRANSMEM 393 413 POTENTIAL.  
 FT TRANSMEM 417 437 POTENTIAL.  
 FT DOMAIN 27 136 HIS-RICH.  
 FT DOMAIN 252 274 HIS-RICH.  
 FT CARBOHYD 337 367  
 FT CONFLICT 367 367  
 FT CONFLICT 403 476  
 SQ SEQUENCE 476 AA; 50656 MW; 48214438BB44919B CRC64;  
 Query Match 16.8%; Score 162; DB 1; Length 476;  
 Best Local Similarity 42.9%; Pred. No. 3.1e-06;  
 Matches 39; Conservative 2; Mismatches 32; Indels 18; Gaps 8;  
 QY 1 HPHKH-HSHQHCHGH-HPAHHPHEDTHQHPHGH-----HPHGH-HPHGHPHGHHPH 53  
 DB 43 HGHSHGHSHDFFHGHSHGHSH-----EDFHGHGHGHSHGHSHGHSHGHSHGHSHGH 97  
 QY 54 GH-HPHCHDQDYPGCDPPPHNQGHCHGHGH 83  
 DB 98 GHSHGHSHDLSLHGH-----GHGHGHSHGH 123  
 RESULT 9  
 DSX\_DROME STANDARD; PRT; 549 AA.  
 ID AC DSX\_DROME STANDARD; PRT; 549 AA.  
 AD P23023; P23022; Q9VHO; G  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 10-NOV-1991 (Rel. 20, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Doublesex protein.  
 GN DSX OR Cg11094.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS FEMALE AND MALE).  
 RC TISSUE=Larva, and Pupae;  
 RX MEDLINE=89168451; PubMed=24933994;  
 RA Burtis K.C., Baker B.S.;  
 RT "Drosophila doublesex gene controls somatic sexual differentiation by  
 RT producing alternatively spliced mRNAs encoding related sex-specific  
 RT polypeptides.";  
 RL Cell 56:997-1010(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM MALE).  
 RC STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Achavani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP DNA-BINDING.  
 RX MEDLINE=91330881; PubMed=1907913;  
 RA Burtis K.C., Coschigano K.T., Baker B.S., Wensink P.C.;  
 RT "The doublesex proteins of Drosophila melanogaster bind directly to a  
 RT sex-specific yolk protein gene enhancer.";  
 RL EMO J. 10:2577-2582(1991).  
 RN [4]  
 RP DNA-BINDING DOMAIN, AND MUTAGENESIS.  
 RX MEDLINE=93178426; PubMed=8440242;  
 RA Erdman S.E., Burtis K.C.;  
 RT "The Drosophila doublesex proteins share a novel zinc finger related  
 RT DNA binding domain.";  
 RL EMO J. 12:527-535(1993).  
 RN [5]  
 RP FUNCTION.  
 RC STRAIN=Canton-S;  
 RX MEDLINE=2233098; PubMed=12435630;  
 RA Dauwalder B., Tsujimoto S., Moss J., Mattox W.;  
 RT "The Drosophila takeout gene is regulated by the somatic sex-  
 RT determination pathway and affects male courtship behavior.";  
 RL Genes Dev. 16:2879-2892(2002).  
 CC -!- FUNCTION: Controls somatic sexual differentiation. Binds directly  
 CC and specifically to the FBE (fat body enhancer) of the yolk  
 CC protein 1 and 2 genes (Yp1 and Yp2). This enhancer is sufficient  
 CC to direct the female-specific transcription characteristic of the  
 CC Yp genes in adult fat bodies. Involved in regulation of male-  
 CC specific expression of takeout in brain-associated fat body.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Male;  
 CC IsoId=P23023-1; Sequence=Displayed;  
 CC Name=Female;  
 CC IsoId=P23023-2; Sequence=VSP\_001321, VSP\_001322;  
 CC -!- MISCELLANEOUS: EXPERIMENTALLY SHOWN TO BIND ZINC.  
 CC -!- SIMILARITY: Contains 1 DM domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M25292; AAA17840.1; -;

DR EMBL; M25293; AAA17841.1; -;  
DR EMBL; M25294; AAA17842.1; -;  
DR EMBL; AE003676; BAF54168.1; -;  
DR PIR; A32372; A32372;  
DR PIR; B32372; B32372;  
DR PDB; 1LPV; 02-OCT-02;  
DR TRANSFAC; T00955; -;  
DR TRANSFAC; T00956; -;  
DR FlyBase; FBgn0000504; dex;  
DR GO; GO:0003729; F:RNA binding; NAS;  
DR GO; GO:0003700; P:transcription factor activity; NAS;  
DR GO; GO:0007619; P:courtship behavior; NAS;  
DR GO; GO:0045497; P:female anatal morphogenesis (sensu Holometabola); NAS;  
DR GO; GO:0007486; P:female genital morphogenesis (sensu Holometabola); NAS;  
DR GO; GO:0009101; P:female somatic sex determination; NAS;  
DR GO; GO:0009101; P:female somatic sex determination; NAS;  
DR GO; GO:0007483; P:genital disc metamorphosis; NAS;  
DR GO; GO:0045496; P:female anatal morphogenesis (sensu Holometabola); NAS;  
DR GO; GO:0045433; P:male courtship behavior (sensu Insecta); so.; NAS;  
DR GO; GO:0007485; P:female genital morphogenesis (sensu Holometabola); NAS;  
DR GO; GO:0019102; P:female somatic sex determination; NAS;  
DR GO; GO:0045892; P:negative regulation of transcription; NAS;  
DR GO; GO:0045893; P:negative regulation of transcription; NAS;  
DR GO; GO:0007548; P:positive regulation of transcription; NAS;  
DR GO; GO:0007548; P:positive regulation of transcription; NAS;  
DR InterPro; IPR001275; DM-DNA-binding;  
DR Pfam; PF00751; DM-domain; 1;  
DR SMART; SM00301; DM; 1;  
DR PROSITE; PS40000; DM DOMAIN 1; 1;  
DR PROSITE; PS50809; DM DOMAIN 2; 1;  
DR Sexual differentiation regulation; Alternative splicing; DNA-binding;  
KW Transcription regulation; Nuclear protein; Zinc; Metal-binding;  
KW 3D-structure.  
FT DNA\_BIND 44 91 DM.  
FT DOMAIN 119 224 SER/GLY-RICH.  
FT DOMAIN 267 296 SER/GLY-RICH.  
FT VARSPLIC 398 427 ARVENRTVAQIVYNYTPMALVNGAPMYL -> GQVYVNE  
FT YSRQNLNIYDGGELRNTTRQCG (in isoform Female).  
FT /FTID=VSP 001321.  
FT Missing (in isoform Female).  
FT VARSPLIC 428 549 /FTID=VSP 001322.  
FT C->A:H: ABOLISHES DNA-BINDING.  
FT C->Y: ABOLISHES DNA-BINDING.  
FT H->Y: ABOLISHES DNA-BINDING.  
FT H->Y: ABOLISHES DNA-BINDING.  
FT C->D.Y: ABOLISHES DNA-BINDING.  
FT C->Y: ABOLISHES DNA-BINDING.  
FT R->Q: ABOLISHES DNA-BINDING.  
FT SEQUENCE 549 AA; 57409 MW; 3C1B92724B4CE083 CRC64;  
Query Match 16.6%; Score 160; DB 1; Length 549;  
Best Local Similarity 28.8%; Pred. No. 5.2e-06;  
Matches 42; Conservative 3; Mismatches 37; Indels 64; Gaps 9;  
QY 1 HPKHSHSEHQPHGHHPHHPHE-HDTHRQ-----H 31  
Db 128 HVYAHVHAHHAHGGH-HSHGHVHLH-HQAAAAAAPSASHLGSSSTAASSIGH 184  
QY 32 PHGHHPH-----GHHPHGHHPHGHHPH-GHHPCHDQDVGPCDP-----PP 72  
Db 185 AAHAHVHMAAAAAAASVAQHQHQSHPHSHHHHQNHHQHPHQ-----QPATQTLRSP 237  
QY 73 HNGCHCHGCHGPPFHLRRGCKGP 98  
Db 238 HSD-----HGGVGATPSSSGGGAP 257  
RESULT 10  
AMV2 SCHPO  
ID AMV2 SCHPO STANDARD; PRT; 774 AA.  
AC Q42918; Q96WR2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Probable alpha-amylase meu7 precursor (EC 3.2.1.1) (1,4-alpha-D-

DE Glucan glucanohydrolase) (Meiotic expression upregulated protein 7).  
GN MEU7 OR SPBC16A3.13.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
RA Scouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
RA Bozym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
RN [2]  
RP SEQUENCE OF 525-774 FROM N.A.  
RC STRAIN=CD16-1;  
RX MEDLINE=21270454; PubMed=11376151;  
RA Watanabe T., Miyashita K., Saito T.T., Yoneki T., Kakiyama Y.,  
RA Nishishima K., Kishi Y.A., Shimoda C., Nojima H.;  
RT "Comprehensive isolation of meiosis-specific genes identifies novel  
RT proteins and unusual non-coding transcripts in Schizosaccharomyces  
RT pombe.";  
RL Nucleic Acids Res. 29:2327-2337(2001).  
CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-alpha-glucosidic  
CC linkages in oligosaccharides and polysaccharides.  
CC -!- COPACTOR: Binds 1 calcium ion per subunit (potential).  
CC -!- SIMILARITY: Belongs to family 13 of glycosyl hydrolases.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AL021748; CAAL6864.1; -;  
CC EMBL; AB054318; BAB60884.1; -;  
CC PIR; T39539; T39539.  
CC HSP; P56271; 2A8A.  
CC GeneDB\_SPombe; SPBC16A3.13; -;  
CC InterPro; IPR006047; Alpha\_amyl\_cat.  
CC Pfam; PF00128; alpha-amylase; 1.  
KW Meiosis; Carbohydrate metabolism; Hydrolase; Glycosidase;  
KW Calcium-binding; Signal; Glycoprotein.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 774 PROBABLE ALPHA-AMYLASE MEU7.  
FT ACT\_SITE 461 461 BY SIMILARITY.  
FT ACT\_SITE 553 553 BY SIMILARITY.

```
FT METAL 143 143 CALCIUM 1 (POTENTIAL).
FT METAL 395 395 CALCIUM 1 (VIA CARBONYL OXYGEN)
FT 52 60 BY SIMILARITY.
FT DISULFID 229 314 BY SIMILARITY.
FT DISULFID 495 539 BY SIMILARITY.
FT DISULFID 690 724 BY SIMILARITY.
FT CARBOHYD 143 143 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 409 409 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 774 AA; 89459 MW; 92CAAB45664269B CRC64;

Query Match 16.1%; Score 155.5; DB 1; Length 774;
Best Local Similarity 26.8%; Pred. No. 1.6e-05;
Matches 53; Conservative 12; Mismatches 52; Indels 81; Gaps 14;

QY 3 HKHSHEQHPHGH-----PH-----AHPHEHDTHQH-----PHG-----HHPH 38
DB 249 HKPWKEHCCHDHPFPPVPHNGTKPDHPKPKWKEHCCHDXFS-----RPPVPHNGTKPDHPK 308
QY 39 GHHPHGH-----PHG-----HHPHGHHPH-----HD-FQYGPDCDPPPHN----- 74
DB 309 KHEECHHGFRPPVPHNGTKPDHPKPKWKEHCCHDXFS-----RPPVPHNGTKPDHPK 363
QY 75 ---QGHCHGCHG---GPPGHLLRRGPGKPPPHFCROIGSVYVLPRLKGEVLPLEAAPP 129
DB 364 WHEECHHGKFLRPPVPHNVTKPDHPKWEHSECH-----HGRFP 403
QY 130 SFPLPKHKLKPDNOFF 147
DB 404 R-PVPH--NGTKPDHPK 418

RESULT 11
KE4_BRARE ID KE4_BRARE STANDARD; PRT; 352 AA.
AC Q9PUB8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Zinc transporter SLC39A7 homolog (Histidine-rich membrane protein Ke4
DE homolog) (Fragment).
GN HKE4 OR KE4.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP Murray B.W., Suelmann H., Klein J.;
RA "Identification of a homolog of the human HKE4 gene in zebrafish.";
RT Submitted (Oct-1999) to the EMBL/GenBank/DBSJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: Belongs to the ZIP transporter (TC 2.A.5) family.
CC KE4/CatSap subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR ZFIN: ZDB-GENE-991110-20; ke4.
DR InterPro: IPR003689; Zn_transpt_Zip.
DR Pfam: PF02535; Zip; 1.
KW Transport; Transmembrane; Glycoprotein.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 215 235 POTENTIAL.
FT TRANSMEM 318 338 POTENTIAL.
FT DOMAIN 24 105 HIS-RICH.
FT DOMAIN 177 217 HIS-RICH.
FT CARBOHYD 311 311 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT NON_TER 352 352
SQ SEQUENCE 352 AA; 37922 MW; C8C8C60F6D2BA8A6 CRC64;

Query Match 15.9%; Score 154; DB 1; Length 352;
Best Local Similarity 42.7%; Pred. No. 1e-05;
Matches 35; Conservative 2; Mismatches 23; Indels 22; Gaps 8;

QY 1 HPKHSH-----EQHPHG-----HH-----PHAHHPHEHDTHQHPHG-HHPGHHPHG 44
DB 24 HSHHHGHGGGCHGSHGKMGHGASKWSAEANLPHAEHHVHDHGHGTHDHA-HDHC 82
QY 45 H-HPHGH-HPHGH---HPHCHD 61
DB 83 HAHSHGDIHDHGHAKHGHAFD 104

RESULT 12
FXB2_MOUSE ID FXB2_MOUSE STANDARD; PRT; 428 AA.
AC Q64733;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Forkhead box protein B2 (Transcription factor FKX-4).
GN FOXB2 OR FKX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97014266; PubMed=8861101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes fkh-4 and fkh-5 defines domains
RT in the central nervous system.";
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RA SFRN=129;
RA MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RT regulated.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Expressed during embryogenesis.
CC -!- SIMILARITY: Contains 1 fork-head domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J., Houtin D., Houston K.A., Howland T.J., Wei M.H., Ibegwag C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.W., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B., Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinartz K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of *Drosophila melanogaster*.";  
Science 287:2185-2195(2000).  
[3]  
REVIEWS.  
[4]  
MEDLINE=22426069; PubMed=12537572;  
MIMRA S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E., Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;  
"Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review".  
Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
[4]  
SEQUENCE OF 497-625 FROM N.A.  
MEDLINE=90046666; PubMed=2573058;  
Kim Y., Nirenberg M.;  
"Drosophila NK-homeobox genes.";  
Proc. Natl. Acad. Sci. U.S.A. 86:7716-7720(1989).  
-I- FUNCTION: May play a role in specifying the identity of particular somatic muscles and neurons of the CNS.  
-I- TISSUE SPECIFICITY: Mesodermal precursor cells of distinct muscles during embryogenesis, a subset of neuronal cells of the CNS and their precursors and also in cells of a small region of the midgut  
-I- DEVELOPMENTAL STAGE: Postgastrulation-stage.  
-I- SIMILARITY: Belongs to the NK-1 homeobox family.  
-I- SIMILARITY: Contains 1 homeobox domain.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; X55393; CAA39067.1; --  
EMBL; AE003735; AAF55901.3; --  
EMBL; M72289; AA2816.1; --  
PIR; A36664; A36664.  
HSP; P14653; L1B72.  
TRANSFAC; T04257; --

DR FlyBase; FBgn0002941; slow.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; NAS.  
 DR GO; GO:0007501; P:muscle cell fate specification; NAS.  
 DR GO; GO:0007521; P:muscle cell fate determination; IEP.  
 DR InterPro; IPR001356; Homeobox.  
 DR Pfam; PF00046; Homeobox; 1.  
 DR PRINTS; PR00024; HOMEBOX.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; Hox; 1.  
 DR PROSITE; PS00027; HOMEBOX 1; 1.  
 DR PROSITE; PS00071; HOMEBOX 2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 FT Repeat. 201 239 HIS-RICH.  
 FT DOMAIN 221 234 7 X 2 AA TANDEM REPEATS OF H-P.  
 FT DOMAIN 221 234 POLY-ALA.  
 FT DOMAIN 477 522 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 536 542 POLY-GLY.  
 FT DNA BIND 545 604 HOMEBOX.  
 SQ SEQUENCE 659 AA; 69955 MW; 5D401F55C4670280 CRC64;  
 Query Match 15.5%; Score 149.5; DB 1; Length 659;  
 Best Local Similarity 30.6%; Pred. No. 4.2e-05;  
 Matches 44; Conservative 4; Mismatches 49; Indels 47; Gaps 9;  
 QY 10 QHPHG---HHPH-AHHPHEDHTROHPGHHPHGHHPHGH-HPHGH-HPHCHDF 62  
 DB 175 QHPHALLQHPHLLQNPQLAAQGHM-HHQHQHQHPHHPHSHQHPHHPH----- 229  
 QY 63 QDYGCPDPPPHNCHCHGCHGPPHLLRRGPKGPRPFHCQIGQSVYVLPPLRKGEVLP 122  
 DB 230 -----PHDPSA-----VFELRAPSSTAPP--SPATSP 257  
 QY 123 LPEANPSPFLPHKHLKEDNOP 146  
 DB 258 LSP---PTSPAMHSDDQMSPIAP 278  
 RESULT 14  
 PYGO DROME  
 ID PYGO DROME STANDARD; PRT; 815 AA.  
 AC QSV9W8;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE PYGopus protein (Gammy legs protein).  
 GN PYGO OR GAM OR CG1518.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 CC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RA Murone M., Zuellig S., Basler K.;  
 RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
 RT pygopus to the nuclear beta-catenin-TCF complex.";  
 RL Cell 109:47-60(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Head;  
 RX MEDLINE=22010046; PubMed=12015286;  
 RA Parker D.S., Jemison J., Cadigan K.M.;  
 RT "Pygopus, a nuclear PHD-finger protein required for wingless signaling  
 RT in Drosophila.";  
 RL Development 129:2565-2576(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=107311132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotaker P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 CC -!- FUNCTION: Involved in signal transduction through the Wnt pathway.  
 CC -!- SUBUNIT: Binds to BCL9 via the PHD-type zinc finger motif, and  
 CC thereby becomes part of the nuclear ARM/PAN complex.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Ubiquitous throughout embryogenesis and larval  
 CC development.  
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 CC throughout development.  
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF457206; AAL91369.1; -  
 DR EMBL; AY075095; AAL79357.1; -  
 DR EMBL; AE003778; AAF57161.1; -  
 DR EMBL; AY058500; AAL13729.1; -  
 DR FlyBase; FBgn0043900; pygo.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0030528; P:transcription regulator activity; IPI.  
 DR GO; GO:0030177; P:positive regulation of Wnt receptor signal. . . ; IPI.

DR GO: GO:0007367; P-segment polarity determination; IMP.  
 DR GO: GO:0016055; P-wnt receptor signaling pathway; IGI.  
 DR InterPro: IPR001955; Znf\_PHD.  
 DR Pfam: PF00628; PHD; 1.  
 DR SMART: SM00249; PHD; 1.  
 DR PROSITE: PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE: PS00016; ZF\_PHD\_2; 1.  
 KW Nuclear protein; trans-acting factor; Wnt signaling pathway; Zinc;  
 KW Nuclear-binding; Zinc-finger; Segmentation polarity protein.  
 FT DOMAIN 39 45  
 FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT ZN\_FING 747 805  
 FT PHD-TYPE.  
 FT DOMAIN 48 65  
 FT ALA-RICH.  
 FT DOMAIN 123 749  
 FT ASN/GLY/HIS/MET/PRO-RICH.  
 FT CONFLICT 393 393  
 FT S -> P (IN REF. 1).  
 SQ SEQUENCE 815 AA; 50493 MW; 369FDSASD34BC136 CRC64;

Query Match 15.5%; Score 149.5; DB 1; Length 815;  
 Best Local Similarity 32.5%; Pred. No. 5.1e-05;  
 Matches 53; Conservative 7; Mismatches 38; Indels 65; Gaps 14;

QY 8 HECH---PHGHHP-----HAHPHEDHTHQPHG---HHPHGHHPGHHPGHHP 52  
 Db 542 HQCHMGGGPGHGGPMGNGMGNQMLPQOPSHLGGPHPNMNH- -HHPH-HHPGGPPPP 598

QY 53 H-----GHHPHCHDFQDYGCPDPPHNOG---HCC---HGH-GPPPGHLRRG-PGKGP 98  
 Db 599 HNMGGPMG-----GPAGPHMGCGPHMGPHMGPHMGPHMGPHMGPHMGPHMG 650

QY 99 RPFHCRQIGSVYRLPLRKGEVLPLPEANFPSP--LPHHKHP 139  
 Db 651 GP-----GGM-----NGPPHPSHPHGHHP 670

RESULT 15  
 ID CBPY SCHPO STANDARD; PRT; 1002 AA.  
 AC O13849; O14366;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carboxypeptidase Y precursor (EC 3.4.16.5) (CPY).  
 GN CPY1 OR PCY1 OR SPAC19G12.10C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A., SUBUNIT, SUBCELLULAR LOCATION, AND MUTAGENESIS OF  
 RP CYS-627.  
 RX MEDLINE=97352672; PubMed=9209031;  
 RA Tabuchi M., Iwahara O., Ohnami Y., Ohuchi N., Sakurai J.-I.,  
 RA Morita T., Iwahara S., Takegawa K.;  
 RT "Vacuolar protein sorting in fission yeast: cloning, biosynthesis,  
 RT transport, and processing of carboxypeptidase Y from  
 RT Schizosaccharomyces pombe".  
 RL J. Bacteriol. 179:4179-4189 (1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders R., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,

RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
 RA Weijens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Bozzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Burnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RT Nature 415:871-880(2002).  
 CC -!- FUNCTION: Involved in degradation of small peptides. Digests  
 CC preferentially peptides containing an aliphatic or hydrophobic  
 CC residue in P1' position, as well as methionine, leucine or  
 CC phenylalanine in P1 position of ester substrate.  
 CC -!- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a  
 CC broad specificity.  
 CC -!- SUBUNIT: Heterodimer of two subunits of 32 kDa and 19 kDa derived  
 CC from the precursor protein and linked by a disulfide bond.  
 CC -!- SUBCELLULAR LOCATION: Lysosome-like vacuoles.  
 CC -!- SIMILARITY: Belongs to peptidase family S10.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D86560; BAA25568.1; -;  
 CC EMBL; Z97209; CAB10121.1; -;  
 DR PIR: T43236; T43236.  
 DR HSP; P00729; LYSC.  
 DR MEROPS; S10.001; -;  
 DR GeneDB SPombe; SPAC19G12.10C; -;  
 DR InterPro; IPR001563; Peptidase S10.  
 DR InterPro; IPR000379; Ser esters  
 DR Pfam; PF00450; serine carboxpept; 1.  
 DR PRINTS; PR00724; CBXOXYPTASEC.  
 DR PRODOM; PD001189; Serine carboxpept; 1.  
 DR PROSITE; PS00131; CARBOXYPEPT SER SER; 1.  
 DR PROSITE; PS00560; CARBOXYPEPT SER HIS; FALSE NEG.  
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 521  
 FT CHAIN 522 1002  
 FT DOMAIN 208 341  
 FT DOMAIN 225 341  
 FT REPEAT 225 237  
 FT REPEAT 238 250  
 FT REPEAT 251 263  
 FT REPEAT 264 276  
 FT REPEAT 277 289  
 FT REPEAT 290 302  
 FT REPEAT 303 315  
 FT REPEAT 316 328  
 FT REPEAT 329 341  
 FT REPEAT 361 423  
 FT REPEAT 361 369  
 FT REPEAT 370 378  
 FT REPEAT 379 387  
 FT REPEAT 388 396  
 FT REPEAT 397 405  
 FT REPEAT 406 414  
 FT REPEAT 415 423  
 FT ACT\_SITE 715 715  
 FT ACT\_SITE 921 921

FT PRO-RICH.  
 FT 9 X 13 AA TANDEM REPEATS OF M-H-H-E-P-G-  
 FT E-H-M-P-P-P-P.  
 FT 1-1.  
 FT 1-2.  
 FT 1-3.  
 FT 1-4.  
 FT 1-5.  
 FT 1-6.  
 FT 1-7.  
 FT 1-8.  
 FT 1-9.  
 FT 7 X 9 AA TANDEM REPEATS OF D-K-E-H-H-K-K-  
 FT G-P-K.  
 FT 2-1.  
 FT 2-2.  
 FT 2-3.  
 FT 2-4.  
 FT 2-5.  
 FT 2-6.  
 FT 2-7 (APPROXIMATE).  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.

Fri Sep 24 09:03:21 2004

```

FT ACT SITE 978 978 BY SIMILARITY.
FT BINDING 924 924 SUBSTRATE (BY SIMILARITY).
FT BINDING 979 979 SUBSTRATE (BY SIMILARITY).
FT DISULFID 627 880 BY SIMILARITY.
FT DISULFID 776 789 BY SIMILARITY.
FT DISULFID 799 822 BY SIMILARITY.
FT DISULFID 806 815 BY SIMILARITY.
FT DISULFID 844 851 BY SIMILARITY.
FT CARBOHYD 659 659 N-LINKED (GLCNAC... ) (POTENTIAL).
FT MUTAGEN 627 627 C->T: 36% OF ORIGINAL ACTIVITY.
SQ SEQUENCE 1002 AA; 114237 MW; 4A8D81CFDAB2D854 CRC64;

Query Match 15.5%; Score 149.5; DB 1; Length 1002;
Best Local Similarity 31.4%; Pred. No. 6.1e-05;
Matches 49; Conservative 5; Mismatches 63; Indels 39; Gaps 10;

QY 2 PHKHSHEQH-----PHGHHPHAHHP-----HEHDTIRQHPHGHPHGHH---PHGHHPHG 49
Db 210 PPMHHRKPGEHMPPPPMHHEFGEHMPPPPMHHEFGEHMPPPPMHHEFGEHMPPPPMHHEFGE 269
QY 50 HH-----PHGHHPHCHDFQDYGPCDPPP--HNOGHCHCHGPPPPGHLRRRGPGKGGRPFHC 103
Db 270 EHMPPPPMHHEFGEH-----NPPPMHHEPGE-----HMPPPPMHHEFGEHMPPPPMH- 317
QY 104 RQIGSVYRLPPL--RKGEVLPLPEANFPSPFLPHHK 137
Db 318 HEFGEHMPPPPMHHEFGEHMPPP-----PFKHHE 346
```

Search completed: September 23, 2004, 22:50:35  
Job time : 20.2369 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 CompuGen Ltd.  
 OM protein - protein search, using sw model  
 Run on: September 23, 2004, 22:25:45 ; Search time 87.3735 Seconds  
 (without alignments)  
 534.449 Million cell updates/sec

Title: US-10-074-225A-5  
 Perfect score: 966  
 Sequence: 1 HPHKHSHEQHPGHHPAH.....PSFPLPHKHLKPDNQPP 148

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_25.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	645.5	66.8	525	11 Q99PS6	Q99PS6 mus musculu
2	645.5	66.8	525	11 Q9ESB3	Q9ESB3 mus musculu
3	645.5	66.8	525	11 Q99PS3	Q99PS3 mus musculu
4	644	66.7	525	11 Q99PS8	Q99PS8 rattus norv
5	644	66.7	546	6 Q9BGU1	Q9BGU1 bos taurus
6	626	64.8	515	11 Q99PS7	Q99PS7 rattus norv
7	614.5	63.6	510	11 Q9ESB2	Q9ESB2 rattus norv
8	281.5	29.1	233	5 Q8MP30	Q8MP30 dictyosteli
9	223.5	23.1	294	2 Q9KI87	Q9KI87 bacillus ce
10	217	22.5	140	5 Q26056	Q26056 plasmodium
11	216.5	22.4	296	2 Q9KI89	Q9KI89 bacillus ce
12	213.5	22.1	465	13 Q801E5	Q801E5 xenopus lae
13	205.5	21.3	273	2 Q9KI88	Q9KI88 bacillus ce
14	205.5	21.3	462	13 Q7ZY91	Q7ZY91 xenopus lae
15	205.5	21.3	462	13 Q7SYH2	Q7SYH2 xenopus lae
16	205	21.2	295	16 Q816T1	Q816T1 bacillus ce

## ALIGNMENTS

RESULT 1

Q99PS6

ID Q99PS6 PRELIMINARY; PRT; 525 AA.

AC Q99PS6; 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Histidine-rich glycoprotein.

GN AW413091 OR MHRG.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/c; TISSUE=Liver;

RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,

RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;

RT "Molecular diversity of mammalian histidine-rich glycoprotein."

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB055897; BAB33094.1; -

DR MGD; MGI:2146636; AW413091.

DR GO; GO:0004859; F:cytosine protease inhibitor activity; IEA.

DR InterPro; IPR000010; Cystatin.

DR Pfam; PF00031; cystatin; 1.

DR SMART; SM00043; Cy; 2.

SQ SEQUENCE 525 AA; 59090 MW; A83B93A439CFB3AC CRC64;

Query Match 66.8%; Score 645.5; DB 11; Length 525;

Best Local Similarity 69.5%; Pred. No. 7.9e-56;

Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPGHHPAHHPHEDTHRQHPGHHPGHHPGHHPGHHPGHHPCH 60

Db 347 HPHSHHPGHSHHGHPGHHPGHSHHGHPGHHPGHHPGHHPGHHPGHHPGH 406

QY 61 DFQYGYGCDPPPHNQ---GHCCHGPPPGHLPBEGGKGRPPHCFQISVVELPLRK 117

Db 407 DFLDYGCDPPSPNSQELKQVHRGYPGPPHSRKRGGKGLFPFPHHQOIGVIVLPLPNT 466

```

QY 118 GEVLPLPEANFPSPFPLPHKHGHPKLPDNOQPP 148
Db 467 GEVLTLPPEANFPSPFSLPNCNRSLOPEIQPPF 497

RESULT 2
Q9ESB3
ID Q9ESB3 PRELIMINARY; PRT; 525 AA.
AC Q9ESB3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-JUN-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN AW413091 OR HRG
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20307726; PubMed=10849117;
RX Hulet M.D., Parish C.R.;
RT "Murine histidine-rich glycoprotein: cloning, characterization and
cellular origin."
RL Immunol. Cell Biol. 78:280-287(2000).
DR EMBL; AF194028; AAG28416.1; -.
DR M3D; MGI:2146636; AW413091.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 525 AA; 59132 MW; 6E55F2A439CFB123 CRC64;

Query Match 66.8%; Score 645.5; DB 11; Length 525;
Best Local Similarity 69.5%; Pred. No. 7.9e-56;
Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 60
Db 347 HPHSHHPGHHSGHHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 406

QY 61 DFQDYGCDPPPNQ---GHCCHGHPGPPGHLRRGPGKGRPFHCRQIGSVYRLPPLK 117
Db 407 DFLDYGCDPPSPNSQELKGQYHRGYGPPHGHGSRKRGKGLFPFHQQIGYVYRLPPLNI 466

QY 118 GEVLPLPEANFPSPFPLPHKHGHPKLPDNOQPP 148
Db 467 GEVLTLPPEANFPSPFSLPNCNRSLOPEIQPPF 497

RESULT 4
Q99PS8
ID Q99PS8 PRELIMINARY; PRT; 525 AA.
AC Q99PS8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein 1.
GN RNRHGI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
Onishi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055895; BAB33092.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match 66.7%; Score 644; DB 11; Length 525;
Best Local Similarity 68.6%; Pred. No. 1.1e-55;
Matches 107; Conservative 7; Mismatches 34; Indels 8; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 55
Db 342 HPHGHHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 401

QY 56 HPHCHDQDYGCDPPPNQ---GHCCHGHPGPPGHLRRRPGKGRPFHCRQIGSVYRL 112
Db 402 HPHGHDFLDYGCDPPSPNSQELKGQYHRGHPHGHGSRKRGKGLFPFHQRQIGYVYRL 461

QY 113 PPLKRGVLPPEANFPSPFPLPHKHGHPKLPDNOQPP 148
Db 462 PPLNVGEVLTLPPEANFPFISLPNCNRSLOPEIQPPF 497

RESULT 5

```

```

RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055898; BAB33095.1; -.
DR EMBL; BC011168; AAH11168.1; -.
DR M3D; MGI:2146636; AW413091.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 525 AA; 59162 MW; A83E93A439CFB126 CRC64;

Query Match 66.8%; Score 645.5; DB 11; Length 525;
Best Local Similarity 69.5%; Pred. No. 7.9e-56;
Matches 105; Conservative 9; Mismatches 34; Indels 3; Gaps 1;

QY 1 HPHKHSHEQHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 60
Db 347 HPHSHHPGHHSGHHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 406

QY 61 DFQDYGCDPPPNQ---GHCCHGHPGPPGHLRRGPGKGRPFHCRQIGSVYRLPPLK 117
Db 407 DFLDYGCDPPSPNSQELKGQYHRGYGPPHGHGSRKRGKGLFPFHQQIGYVYRLPPLNI 466

QY 118 GEVLPLPEANFPSPFPLPHKHGHPKLPDNOQPP 148
Db 467 GEVLTLPPEANFPSPFSLPNCNRSLOPEIQPPF 497

RESULT 4
Q99PS8
ID Q99PS8 PRELIMINARY; PRT; 525 AA.
AC Q99PS8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein 1.
GN RNRHGI.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
Onishi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055895; BAB33092.1; -.
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 525 AA; 59049 MW; 38290A631FAC7777 CRC64;

Query Match 66.7%; Score 644; DB 11; Length 525;
Best Local Similarity 68.6%; Pred. No. 1.1e-55;
Matches 107; Conservative 7; Mismatches 34; Indels 8; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 55
Db 342 HPHGHHPGHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 401

QY 56 HPHCHDQDYGCDPPPNQ---GHCCHGHPGPPGHLRRRPGKGRPFHCRQIGSVYRL 112
Db 402 HPHGHDFLDYGCDPPSPNSQELKGQYHRGHPHGHGSRKRGKGLFPFHQRQIGYVYRL 461

QY 113 PPLKRGVLPPEANFPSPFPLPHKHGHPKLPDNOQPP 148
Db 462 PPLNVGEVLTLPPEANFPFISLPNCNRSLOPEIQPPF 497

RESULT 5

```

```

Q9BGU1
ID Q9BGU1 PRELIMINARY; PRT; 546 AA.
AC Q9BGU1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN BTHRG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055894; BAB33093.1; -
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 66.7%; Score 644; DB 6; Length 546;
Best Local Similarity 62.7%; Pred. No. 1.1e-55;
Matches 104; Conservative 10; Mismatches 34; Indels 18; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAEHDTROHPGHHPGHHPGHHPGHHPGHHPGH 60
DB 353 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 412
QY 61 -----DFQYGCDDPPPHN---QGHCHGHGPPPGHLRRRGKGRPF 102
DB 413 HPGYHHPGHHPHNDYDHGCDPPPHRQDPQDHRHQGRDPPRSHKRGKGRPF 472
QY 103 CRQIGSVYLPRLKGEVLPLEANFPPSPPLPHKHPKLPDNP 148
DB 473 WRPTGYIHLPLSLKGEVLPLEANFPPSPPLPHKHPKLPDNP 518

RESULT 6
Q9PS7
ID Q9PS7 PRELIMINARY; PRT; 515 AA.
AC Q9PS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein 2.
GN BTHRG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055896; BAB33093.1; -
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 64.8%; Score 626; DB 11; Length 515;
Best Local Similarity 69.5%; Pred. No. 6.5e-54;
Matches 105; Conservative 6; Mismatches 32; Indels 8; Gaps 2;

Q9BGU1
ID Q9BGU1 PRELIMINARY; PRT; 546 AA.
AC Q9BGU1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN BTHRG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055894; BAB33093.1; -
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 546 AA; 61948 MW; 26264858824D89EE CRC64;

Query Match 66.7%; Score 644; DB 6; Length 546;
Best Local Similarity 62.7%; Pred. No. 1.1e-55;
Matches 104; Conservative 10; Mismatches 34; Indels 18; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAEHDTROHPGHHPGHHPGHHPGHHPGHHPGH 60
DB 353 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 412
QY 61 -----DFQYGCDDPPPHN---QGHCHGHGPPPGHLRRRGKGRPF 102
DB 413 HPGYHHPGHHPHNDYDHGCDPPPHRQDPQDHRHQGRDPPRSHKRGKGRPF 472
QY 103 CRQIGSVYLPRLKGEVLPLEANFPPSPPLPHKHPKLPDNP 148
DB 473 WRPTGYIHLPLSLKGEVLPLEANFPPSPPLPHKHPKLPDNP 518

RESULT 6
Q9PS7
ID Q9PS7 PRELIMINARY; PRT; 515 AA.
AC Q9PS7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein 2.
GN BTHRG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Wakabayashi S., Takahashi K., Hirokado Y., Togo Y., Izumi S.,
RA Ohashi T., Sato N., Hirata D., Tsuchida N., Koide T.;
RT "Molecular diversity of mammalian histidine-rich glycoprotein.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055896; BAB33093.1; -
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 515 AA; 58055 MW; 7CEBA3A1A3678966 CRC64;

Query Match 64.8%; Score 626; DB 11; Length 515;
Best Local Similarity 69.5%; Pred. No. 6.5e-54;
Matches 105; Conservative 6; Mismatches 32; Indels 8; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAEHDTROHPGHHPGHHPGHHPGHHPGHHPGH 60
DB 342 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 396
QY 61 DFQYGCDDPPPHN---GHCHGHGPPPGHLRRRGKGRPFPHCRQIGSVYLP 117
DB 397 DFQYGCDDPPPSNSQELKQYHGHGHPGHGSRKRGKGLFPFHQRQIGVYV 456
QY 118 GEVLPLPEANFPPSPPLPHKHPKLPDNP 148
DB 457 GEVLPLPEANFPPSPPLPHKHPKLPDNP 148
QY 118 GEVLPLPEANFPPSPPLPHKHPKLPDNP 148
DB 457 GEVLPLPEANFPPSPPLPHKHPKLPDNP 148

RESULT 7
Q9SEB2
ID Q9SEB2 PRELIMINARY; PRT; 510 AA.
AC Q9SEB2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Histidine-rich glycoprotein.
GN HRG.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Lewis;
RA Hulet M.D., Parish C.R.;
RT "Marine histidine-rich glycoprotein: cloning, characterization and
RT cellular origin.";
RL Immunol. Cell Biol. 78:280-287(2000).
DR EMBL; AF194029; AAG28417.1; -
DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 2.
DR SEQUENCE 510 AA; 57581 MW; 508B6E06AA2ED58E CRC64;

Query Match 63.6%; Score 614.5; DB 11; Length 510;
Best Local Similarity 68.9%; Pred. No. 8.9e-53;
Matches 102; Conservative 7; Mismatches 32; Indels 7; Gaps 2;

QY 1 HPHKHSHEQHPGHHPHAEHDTROHPGHHPGHHPGHHPGHHPGHHPGH 60
DB 342 HPHGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGHHPGH 396
QY 61 DFQYGCDDPPPHN---GHCHGHGPPPGHLRRRGKGRPFPHCRQIGSVYLP 120
DB 397 DFQYGCDDPPPSNSQYH---QGHGPPHGHGSRKRGKGLFPFHQRQIGVYV 454
QY 121 LPLPEANFPPSPPLPHKHPKLPDNP 148
DB 455 LTPPEANFPPSPPLPHKHPKLPDNP 148

RESULT 8
Q9MP30
ID Q9MP30 PRELIMINARY; PRT; 233 AA.
AC Q9MP30; Q8T164;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (Similar to plasmodium lophurae histidine-rich
DE glycoprotein).
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.

```

RC	STRAIN=AX4;
RA	Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA	Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA	Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT	"Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RL	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN=AX4;
RC	MEDLINE=22092622; PubMed=12097910;
RX	Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA	Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA	Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT	"Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL	Nature 418:79-85(2002).
RL	[3]
RN	SEQUENCE FROM N.A.
RP	STRAIN=AX4;
RC	Baumgart C.;
RL	Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC123513; AAM44363.1; -
DR	WHL; AC117070; AAM09303.2; -
KW	Hypothetical protein.
SQ	SEQUENCE 233 AA; 28937 MW; AD4C7A2F86B3B14B9 CRC64;
Query Match	29.1%; Score 281.5; DB 5; Length 233;
Best Local Similarity	38.3%; Pred.No.3.e-20;
Matches	57; Conservative 3; Mismatches 62; Indels 27; Gaps 5;
QY	1 HPHKHSHEQHGHHGHHHAHHPHENDTHRQHFGHHHPGHHPGHHPGHHPGHCH 60
Db	79 HHHHHHHHHHHHHHHHHHPHHHPHHHHHHHHHHHHHHHHHHHHHHHHHHHHH 138
QY	61 DFODYGCDPPPNQGHCCHGGFPPLRRRGKGPRFHCRQIGSVYRLPPLKGEV 120
Db	139 HHH-----HHHHHHHHHHHHHHHHHPHHHPHPHP-----HLHP----- 173
QY	121 LPLPEANFPSPFLPH-HKHGLKPDNQPF 148
Db	174 NPHPHPHPHPHPHPHPHP-NPNHPHP 201
RESULT 9	
O9KI87	PRELIMINARY; PRT; 294 AA.
ID	O9KI87
AC	O9KI87;
DT	01-OCT-2000 (TrEMBLrel. 15, Created)
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Putative VirB (Fragment).
OS	VRRB.
GN	Bacillus cereus.
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX	NCBI_TaxID=1396;
RN	[1]
RN	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 43881;
RC	MEDLINE=20327579; PubMed=10869077;
RX	Schupp J.M., Klevtska A.M., Zinser G., Price L.B., Keim P.;
RT	"VirB, a hypervariable open reading frame in bacillus anthracis.";
RL	J. Bacteriol. 182:3989-3997(2000).
RL	EMBL; AF238886; AAF86203.1; -
FT	NON TER 1
SQ	SEQUENCE 294 AA; 30753 MW; C05B7B37FE2E395C CRC64;
Query Match	23.1%; Score 223.5; DB 2; Length 294;
Best Local Similarity	38.3%; Pred.No.2.e-14;
Matches	49; Conservative 5; Mismatches 41; Indels 33; Gaps 6;
QY	1 HPHKHSHEQHGHHGHHHAHHPHENDTHRQH----PHGHHHPGH--PHG---HPHG 49
Db	118 HGHHGHHHHQGCGHGHGHHHHHQGHGHHHHHHHQGHGHHHHHHHQGHGHHHH 177



Db 118 HHGHHHHHGH-HGHHHHHHHGHGHHHHHGHGHHHHHGHGHHHHHGHG 176  
 QY 48 -----HGHH-----PHGHHPHCHDFQDYPCDPPPHNQHCCH-GHG 83  
 Db 177 QGHQQVHHHHHHHPQAVLYQYQGQGHGHHHGHGQYHG-----HQGHHHHGHQ 229  
 QY 84 PPPGHLRRRGPGKG 97  
 Db 230 QGHGHHGHGQHHG 243  
 RESULT 12  
 Q801E5 PRELIMINARY; PRT; 465 AA.  
 ID Q801E5  
 AC Q801E5  
 DT 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical histidine-rich protein (Fragment).  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22480013; PubMed=12591597;  
 RA Chen Y., Jurgens K., Hollmann T., Claussen M., Ramadori G.,  
 RA Pieler T.  
 RT "Cell-autonomous and signal-dependent expression of liver and  
 RT intestine marker genes in pluripotent precursor cells from Xenopus  
 RT embryos.";  
 RL Mech. Dev. 120:277-298(2003).  
 DR EMBL; AY189284; AAC31610.1; -.  
 DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 2.  
 DR SMART; SM00043; Cy; 2.  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 465 AA; 53528 MW; 0B403AB4F78BBFD4 CRC64;  
 Query Match 22.1%; Score 213.5; DB 13; Length 465;  
 Best Local Similarity 33.1%; Pred. No. 3.5e-13;  
 Matches 50; Conservative 12; Mismatches 64; Indels 25; Gaps 6;  
 QY 1 HPHKHH-SHEQHPGHHPHAPHHEHDTHQHPGHHPGHHPGHHPGHHPGHHPHC 59  
 Db 315 HKRHHPHSHKHKGRHHHHHPHHD-----HPHHHHHHHHHPNHTSSEHHEH 366  
 QY 60 HDFQDYGCDPPPHNQHCCHG-HG-PHPGHLRRRGPGKGRP-FHCROIGSVYRLPLRK 117  
 Db 367 H-----HHPNHTSSSEHSSSEHTDKKAGKADKCFMTRSGVVKITLASE 414  
 QY 118 GEVLPPLPEANF--PSFPLPHKHPKLPDNPQ 146  
 Db 415 TDVLPAPTITISRPSRTEVIQFPEAAASHLP 445  
 RESULT 13  
 Q9KI88 PRELIMINARY; PRT; 273 AA.  
 ID Q9KI88  
 AC Q9KI88  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Putative VrrB (Fragment).  
 OS Bacillus cereus.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1396;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 31293;  
 RX MEDLINE=20327579; PubMed=10869077;  
 RA Schupp J.M., Klevytska A.M., Zinser G., Price L.B., Keim P.;  
 RT "vrrB, a hypervariable open reading frame in bacillus anthracis.";  
 RL J. Bacteriol. 182:3989-3997(2000).  
 DR EMBL; AF233887; AAF86202.1; -.  
 FT NON\_TER  
 SQ SEQUENCE 273 AA; 28573 MW; B80F2979612BEB0C CRC64;  
 Query Match 21.3%; Score 205.5; DB 2; Length 273;  
 Best Local Similarity 36.9%; Pred. No. 1.3e-12;  
 Matches 45; Conservative 3; Mismatches 27; Indels 47; Gaps 7;  
 QY 3 HKHSHSQHPGHHPHAPHHEHDTHQHPH-GHH-----PHGHH-----PHGH 45  
 Db 119 HGHHHHHGHGHHGHHHHGHHGHHHHGHHHHGHHHHGHHHHGHHGHHGHHG 178  
 QY 46 HPHGHH-----PHGHHPHCHDFQDYPCDPPPHNQHCCHGHPGPHHLR 91  
 Db 179 HHGHHHHHPQAVLYQYQGQGHGHH-----HHQGH--HHHQHGHGHHG 222  
 QY 92 RG 93  
 Db 223 QG 224  
 RESULT 14  
 Q7ZY91 PRELIMINARY; PRT; 462 AA.  
 ID Q7ZY91  
 AC Q7ZY91  
 DT 01-JUN-2003 (Tremblrel. 24, Created)  
 DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Similar to fetuin B.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo;  
 RA Klein S., Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC043891; AAF43891.1; -.  
 DR GO; GO:0004869; F:cysteine protease inhibitor activity; IEA.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 2.  
 DR SMART; SM00043; Cy; 2.  
 SQ SEQUENCE 462 AA; 53185 MW; D7BAD339961739FB CRC64;  
 Query Match 21.3%; Score 205.5; DB 13; Length 462;  
 Best Local Similarity 32.5%; Pred. No. 2.2e-12;  
 Matches 49; Conservative 12; Mismatches 65; Indels 25; Gaps 6;  
 QY 1 HPHKHH-SHEQHPGHHPHAPHHEHDTHQHPGHHPGHHPGHHPGHHPGHHPHC 59  
 Db 312 HKRHHPHSHKHKGRHHHHHPHHD-----HPPHHHHHHHHHPNHTSSEHHEH 363  
 QY 60 HDFQDYGCDPPPHNQHCCHG-HG-PHPGHLRRRGPGKGRP-FHCROIGSVYRLPLRK 117  
 Db 364 H-----HHPNHTSSSEHSSSEHTDKKAGKADKCFMTRSGVVKITLASE 411  
 QY 118 GEVLPPLPEANF--PSFPLPHKHPKLPDNPQ 146  
 Db 412 TDVLPAPTITISRPSRTEVIQFPEAAASHLP 442  
 RESULT 15  
 Q7SYH2 PRELIMINARY; PRT; 462 AA.  
 ID Q7SYH2  
 AC Q7SYH2;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Cystatin domain fetuin-like protein.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ventral midgut;  
RA Costa R.M.B., Mason J., Lee M., Amaya E., Zorn A.M.;  
RT "Novel gene expression domains reveal early patterning of the Xenopus  
endoderm";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY260732; AAP82289.1; -.  
SQ SEQUENCE 462 AA; 53186 MW; 796F92774CC27721 CRC64;  
  
Query Match 21.3%; Score 205.5; DB 13; Length 462;  
Best Local Similarity 32.5%; Pred. No. 2.2e-12;  
Matches 49; Conservative 12; Mismatches 65; Indels 25; Gaps 6;  
  
QY 1 HPKHH-SHEQHPGHHHAHPHEHDTHQHPGHHGHPGHHGHPGHHGHPHC 59  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 312  
312 HKRHHPSHKHGKGRHHHPHHD-----HPPHHHHHHHPNHTSSEHHEH 363  
  
QY 60 HDFQDYGPCDPPHNOGHCCHG-PPGHLRRRGKGRP-FHCRQIGSVYRLPLRK 117  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 364  
364 H-----HPPNHTSSGSSSEHTDKKADKCFMTRSKGVVQKITLASE 411  
  
QY 118 GEVLPLPEANF--PSFPLPHKHPKPDNQ 146  
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 412  
412 TDVLPAPTITISRPSPRTEYIQFPEAAHLP 442

Search completed: September 23, 2004, 22:55:03  
Job time : 90.3735 secs

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:47:25 ; Search time 26.747 Seconds  
(without alignments)  
285.664 Million cell updates/sec

Title: US-10-074-225A-5  
Perfect score: 966  
Sequence: 1 HPHKHSHEQHPGHHPHAA.....PSFPLPHKHPKPDNQPPF 148

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/2/aa/5A\_COMB.pep:\*

2: /cgn2\_6/prodata/2/aa/5B\_COMB.pep:\*

3: /cgn2\_6/prodata/2/aa/6A\_COMB.pep:\*

4: /cgn2\_6/prodata/2/aa/8B\_COMB.pep:\*

5: /cgn2\_6/prodata/2/aa/PCFUS\_COMB.pep:\*

6: /cgn2\_6/prodata/2/aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	949	98.2	525	US-09-976-594-64	Sequence 64, Appl
2	185	19.2	26	US-08-785-636-1	Sequence 1, Appl
3	185	19.2	26	US-09-095-407-2	Sequence 2, Appl
4	185	19.2	26	US-09-742-494-2	Sequence 2, Appl
5	184.5	19.1	167	US-09-507-323B-13	Sequence 13, Appl
6	172.5	17.9	400	US-09-543-681A-6151	Sequence 6151, Ap
7	161.5	16.7	225	US-09-252-991A-28819	Sequence 28819, A
8	158.5	16.4	218	US-09-252-991A-25291	Sequence 25291, A
9	155.5	16.1	865	US-09-281-766-19	Sequence 19, Appl
10	148.5	15.4	89	US-09-507-323B-12	Sequence 12, Appl
11	147	15.2	339	US-09-328-352-6551	Sequence 6551, Ap
12	144.5	15.0	60	US-08-255-457-1	Sequence 1, Appl
13	144.5	15.0	60	US-09-115-032-1	Sequence 1, Appl
14	144.5	15.0	60	PCT-US95-05772-1	Sequence 1, Appl
15	140	14.5	189	US-08-152-922A-6	Sequence 6, Appl
16	140	14.5	309	US-08-161-406-2	Sequence 2, Appl
17	135.5	14.0	297	US-09-489-039A-12802	Sequence 12802, A
18	134	13.9	398	US-09-461-474-17	Sequence 17, Appl
19	132	13.7	313	US-08-686-528A-3	Sequence 3, Appl
20	132	13.7	313	US-09-456-287-3	Sequence 3, Appl
21	132	13.7	337	US-08-686-528A-2	Sequence 2, Appl
22	132	13.7	337	US-09-456-287-2	Sequence 2, Appl
23	130	13.5	413	US-09-252-991A-28004	Sequence 28004, A
24	128.5	13.3	474	US-09-461-474-10	Sequence 10, Appl
25	128.5	13.3	613	US-09-252-991A-24982	Sequence 24982, A
26	128.5	13.3	754	US-09-252-991A-22309	Sequence 22309, A
27	126.5	13.1	508	US-09-252-991A-25549	Sequence 25549, A

28 125 12.9 349 3 US-09-461-474-12 Sequence 12, Appl  
29 124 12.8 363 4 US-09-328-352-4930 Sequence 4930, Ap  
30 123.5 12.8 261 4 US-09-602-565-34 Sequence 34, Appl  
31 123.5 12.8 289 4 US-09-540-236-2019 Sequence 2019, Ap  
32 123.5 12.8 448 3 US-09-461-474-8 Sequence 8, Appl  
33 123.5 12.8 630 4 US-09-252-991A-26324 Sequence 26324, A  
34 120.5 12.5 226 4 US-09-252-991A-22052 Sequence 22052, A  
35 120.5 12.5 466 4 US-09-252-991A-26545 Sequence 26545, A  
36 119.5 12.4 277 4 US-08-352-991A-19317 Sequence 19317, A  
37 119 12.3 342 4 US-08-352-991A-26341 Sequence 26341, A  
38 116.5 12.1 591 4 US-09-252-991A-28760 Sequence 28760, A  
39 116 12.0 316 4 US-09-252-991A-27084 Sequence 27084, A  
40 115.5 12.0 294 4 US-09-252-991A-24096 Sequence 24096, A  
41 115 11.9 380 4 US-09-252-991A-22319 Sequence 22319, A  
42 114.5 11.9 274 4 US-09-711-164-369 Sequence 369, App  
43 114.5 11.3 274 4 US-09-711-164-407 Sequence 407, App  
44 114 11.8 480 1 US-07-882-292-2 Sequence 2, Appl  
45 114 11.8 480 2 US-08-331-644-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-09-976-594-64  
; Sequence 64, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Suchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 64  
; LENGTH: 525  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. 6673549 085596CD1  
US-09-976-594-64

Query Match 98.2%; Score 949; DB 4; Length 525;  
Best Local Similarity 98.6%; Pred. No. 1.3e-85;  
Matches 146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 HPHKHSHEQHPGHHPHAPHEDTHRQHPGHHPGHHPGHHPGHHPGHHPCH 60  
Db 350 HPHKHSHEQHPGHHPHAPHEDTHRQHPGHHPGHHPGHHPGHHPGHHPCH 409  
QY 61 DFQDYGCDPPHNOGCHGHPGPPGHLRRRGKGPGRPFPCQIGSVYRLPKGEV 120  
Db 410 DFQDYGCDPPHNOGCHGHPGPPGHLRRRGKGPGRPFPCQIGSVYRLPKGEV 469  
QY 121 LPLPEANFPSPFLPHHHPKHPKPDNQPPF 148  
Db 470 LPLPEANFPSPFLPHHHPKHPKPDNQPPF 497

RESULT 2  
US-08-785-636-1  
; Sequence 1, Application US/08785636  
; Patent No. 6027942  
; GENERAL INFORMATION:  
; APPLICANT: Yip, Tai-Tung  
; APPLICANT: Hutchens, T. William  
; TITLE OF INVENTION: Method and Apparatus for Desorption and Ionization of  
; TITLE OF INVENTION: Analyses

```
; FILE REFERENCE: Hutchens
; CURRENT APPLICATION NUMBER: US/08/785,636
; CURRENT FILING DATE: 1997-01-17
; EARLIER APPLICATION NUMBER: 08/068,896
; EARLIER FILING DATE: 1993-05-28
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: peptide
US-08-785-636-1

Query Match      19.2%; Score 185; DB 3; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GHHPHGHHPGHHPGHHPGHHPGHHP 58
Db      1 GHHPHGHHPGHHPGHHPGHHPGHHP 25

RESULT 4
US-09-742-494-2
; Sequence 2, Application US/09742494
; Patent No. 6528320
; GENERAL INFORMATION:
; APPLICANT: HUTCHENS, WILLIAM
; APPLICANT: YIP, TAI-TUNG
; TITLE OF INVENTION: METHODS AND APPARATUS FOR DESORPTION AND IONIZATION OF ANALYTES
; FILE REFERENCE: P00795USE/09306611
; CURRENT APPLICATION NUMBER: US/09/742,494
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 2
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-742-494-2

Query Match      19.2%; Score 185; DB 4; Length 26;
Best Local Similarity 100.0%; Pred. No. 6.9e-12;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      34 GHHPHGHHPGHHPGHHPGHHPGHHP 58
Db      1 GHHPHGHHPGHHPGHHPGHHPGHHP 25

RESULT 5
US-09-507-323B-13
; Sequence 13, Application US/09507323B
; Patent No. 6274345
; GENERAL INFORMATION:
; APPLICANT: Choi, Jong Hyun
; APPLICANT: Lee, Sang Yup
; APPLICANT: Xu, Zhaohui
; TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE
; TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E. COLI AS CELL SURFACE ANCHORING
; TITLE OF INVENTION: MOTIF
; FILE REFERENCE: HYLEE39.001AUS
; CURRENT APPLICATION NUMBER: US/09/507,323B
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: KR 10-1999-0005773
; PRIOR FILING DATE: 1999-02-22
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Histidine linker.
US-09-507-323B-13

Query Match      19.1%; Score 184.5; DB 3; Length 167;
Best Local Similarity 32.7%; Pred. No. 5.6e-11;
Matches 48; Conservative 2; Mismatches 50; Indels 47; Gaps 9;

QY      1 HPHKHS-----HEQH-----PHCH-----PHAHHPHHD-----THQH--- 31
Db      8 HHHHHSGLDPSGHHHHHHSGLDPSGHHHHHHSGLDPSGHHHHHHSGLDPSGHHHHHHS 67
QY      32 --PHGHHPHGH---HPHGHHPHGH---HPHGHHPHGHFDQ-----YGCDDPPPHNQ 75
```

```

Query Match          16.7%  Score 161.5;  DB 4;  Length 225;
Best Local Similarity 32.3%  Pred. No. 1.4e-08;
Matches             52;  Conservative 3;  Mismatches 49;  Indels 57;  Gaps 11;

```

RESULT 9  
US-09-281-766-19  
; Sequence 19, Application US/09281766  
; Patent No. 6376196  
; GENERAL INFORMATION:  
; APPLICANT: Conrad, Patricia C.  
; APPLICANT: Louis, Kitland  
; APPLICANT: The Regents of the University of California  
; TITLE OF INVENTION: Recombinant Neospora Antigens and Their Uses  
; FILE REFERENCE: 023070-082510US  
; CURRENT APPLICATION NUMBER: US/09/281,766  
; CURRENT FILING DATE: 1999-03-30  
; PRIOR APPLICATION NUMBER: US 08/645,951  
; PRIOR FILING DATE: 1996-05-10  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 865  
; TYPE: PRT  
; ORGANISM: Neospora sp.  
; OTHER INFORMATION: NC-p65 cdNA  
US-09-281-766-19

Query Match 16.1%; Score 155.5; DB 4; Length 865;  
 Best Local Similarity 29.2%; Pred. No. 2.3e-07;  
 Matches 47; Conservative 9; Mismatches 88; Indels 17; Gaps 6;

QY 2 PHKHSHHQPHGHHPHHEHDTROHPHGHHPHGHHPHGHHPHGHHPHCHD 61  
 DB 571 PDQCHPHPRNPPEASPPSPNPNQHPHPPHPPNPPEASPPSPNPNQHPHPP 630  
 QY 62 FQYGP-CDPP-PPHNGHCHGHGPPPPHGLRERGPKGP-PPF-HCRQIGSV 109  
 DB 631 PRNPCCASPPSPNPNQHPHPPHPPNPPEASPPSPNPNQHPHPPHPPH 689  
 QY 110 YRLP-PP-PLKGEVLPPEANFSPPLPHKHHPKLPKPNQ 146  
 DB 690 HPYGYNLPYTYHQSLPYGPYGRDPCPCASHPYPADDS 730

RESULT 10  
 US-09-507-323B-12  
 ; Sequence 12, Application US/09507323B  
 ; Patent No. 6274345  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Choi, Jong Hyun  
 ; APPLICANT: Lee, Sang Yup  
 ; APPLICANT: Xu, Zhaohui  
 ; TITLE OF INVENTION: EXPRESSION VECTORS COMPRISING A GENE  
 ; TITLE OF INVENTION: ENCODING OMPC DERIVED FROM E. COLI AS CELL SURFACE ANCHORING  
 ; TITLE OF INVENTION: MOTIF  
 ; FILE REFERENCE: HYLEE39.001AUS  
 ; CURRENT APPLICATION NUMBER: US/09/507,323B  
 ; CURRENT FILING DATE: 2000-02-18  
 ; PRIOR APPLICATION NUMBER: KR 10-1999-0005773  
 ; PRIOR FILING DATE: 1999-02-22  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 89  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Histidine linker.  
 US-09-507-323B-12

Query Match 15.4%; Score 148.5; DB 3; Length 89;  
 Best Local Similarity 42.5%; Pred. No. 9.9e-08;  
 Matches 31; Conservative 0; Mismatches 27; Indels 15; Gaps 4;

QY 1 HPHKHS-----HQHPHGHHPHHPHHEHDTROHPHGHHPHCHD 50  
 DB 8 HHHHHHGLDPSGHHHHHGLDPSGHHHHH--GLDPSGHHHHHGLDPSGHHHHH 65  
 QY 51 ---HPHGHHPHCH 60  
 DB 66 SGLDPSGHHHHH 78

RESULT 11  
 US-09-328-352-6551  
 ; Sequence 6551, Application US/09328352  
 ; Patent No. 6562958  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary L. Breton et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
 ; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC99-03PA  
 ; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6551  
 ; LENGTH: 339  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6551

Query Match 15.2%; Score 147; DB 4; Length 339;  
 Best Local Similarity 27.8%; Pred. No. 5.7e-07;  
 Matches 47; Conservative 4; Mismatches 28; Indels 90; Gaps 12;

QY 3 KKHSHHQH-----PHGHHPHHPH-----PHEDTTHQHPHG 34  
 DB 132 HGHHLRHHHNGCGHGLDLNGLHGHHLRHHHNGCGHGLHGLHGHHLRHH-HGCGGH 190  
 QY 35 -----HHPHGHHPHG-----HHPHGHHPHG-----HHPHC 59  
 DB 191 DFGLSLDRDLHLRHH-HGCGHGLDLNGLHGHHLRHH-HGCGHGLHGLHGHHLR 248  
 QY 60 HFDQYGFCDPPPHNQGH-----CCHGH-----GPPFGLHRR 91  
 DB 249 H-----HHGCD---HGLGLSLDRDLHLRHHHDCDHGHLRSLDRDLHLR 290

RESULT 12  
 US-08-255-457-1  
 ; Sequence 1, Application US/08255457  
 ; Patent No. 5780040  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Plaut, Andrew G.  
 ; APPLICANT: Gilbert-Rothstein, Joanne V.  
 ; APPLICANT: Wright, Andrew  
 ; TITLE OF INVENTION: HELICOBACTER PYLORI NICKEL BINDING  
 ; TITLE OF INVENTION: PROTEIN  
 ; NUMBER OF SEQUENCES: 3  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Fish & Richardson  
 ; STREET: 225 Franklin Street  
 ; CITY: Boston  
 ; STATE: Massachusetts  
 ; COUNTRY: U.S.A.  
 ; ZIP: 02110-2804  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/255,457  
 ; FILING DATE:  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Clark, Paul C.  
 ; REGISTRATION NUMBER: 30,162  
 ; REFERENCE/DOCKET NUMBER: 00398/090001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 542-5070  
 ; TELEFAX: (617) 542-8906  
 ; TELEX: 200154  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 60 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-255-457-1

Query Match 15.0%; Score 144.5; DB 1; Length 60;  
 Best Local Similarity 40.9%; Pred. No. 1.6e-07;  
 Matches 27; Conservative 3; Mismatches 25; Indels 11; Gaps 3;

QY 19 AHHPHEDTTHQHPHGHHPHGHHPHGHHPHGHHPHGHHPHCHDFOYGFCDPPP-HNQG 76  
 DB 2 AHHEEQHGH-HHHHHHTHHYHGEHHHHHSSHH-----EEGCSSTDSHHQEE 52  
 QY 77 HCCHGH 82  
 DB 53 GCCHGH 58

```

CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05772
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul C.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00398/090001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-05772-1

Query Match          15.0%; Score 144.5; DB 5; Length 60;
Best Local Similarity 40.9%; Pred. No. 1.6e-07;
Matches 27; Conservative 3; Mismatches 25; Indels 11; Gaps 3;

QY      19 AHHPEHDTHRCHPHGHHPHGHHHPGHHHPGHCHDFODYGPCDPPP--HNQG 76
       |||||
Db       2 AHHEQRGGH--HHHHHTHHHYGGRHHHHHSHH-----EGCCSTSDSHHQEE 52
       |||||

QY      77 HCHGH 82
       |||||
Db       53 GCCHGH 58

RESULT 15
US-08-152-922A-6
Sequence 6, Application US/08152922A
Patent No. 5395614
GENERAL INFORMATION:
APPLICANT: Knapp,, Bernhard
APPLICANT: Hundt,, Erika
APPLICANT: Enders, Burkhard
APPLICANT: Kuepper, Hans
TITLE OF INVENTION: Protective Plasmodium Falciparum
TITLE OF INVENTION: Hybrid Proteins which Contain Part-Sequences of the Malaria
TITLE OF INVENTION: Antigenis HRPII and SERP, the Preparation and Use Thereof
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W. Suite 700
CITY: Washington
STATE: D.C.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/152,922A
FILING DATE: 16-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/806,471

```

```

; FILING DATE: 13-DEC-1991
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481-1143-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 189 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-152-922A-6

Query Match      14.5%; Score 140; DB 1; Length 189;
Best Local Similarity 33.0%; Pred. No. 1.5e-06;
Matches 37; Conservative 5; Mismatches 34; Indels 36; Gaps 11;

QY      3 HKHSHSHOHPHGHHP-----HANHP-----HEHDTHQHPGHHP-----HGHP-----H 43
DB      7 HAHHAADAH HAHHAADAHAAHHAAYAHHAHAADAHAH HAHHAADAHHAHAADAHH 64

QY      44 GHHP-----HGHP-----HCHDFQD-YGPCDPPPHNQGHCHGH 82
DB      65 AHHAADAHHAHAADAHHAHAADAHHAHAASDAHHAAD--AHHAAYAHHAH 114

```

Search completed: September 23, 2004, 22:56:48  
Job time : 27.747 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:55:11 ; Search time 101.044 Seconds  
(without alignments)  
470.989 Million cell updates/sec

Title: US-10-074-225A-5  
Perfect score: 966  
Sequence: 1 HPHKHSHEQHPGHGHPAH.....PSFPLPHKHPKPDNOFPF 148

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1349238 seqs, 321558718 residues

Total number of hits satisfying chosen parameters: 1349238

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	966	100.0	148	14	US-10-074-225A-5
2	966	100.0	525	14	US-10-074-225A-1
3	949	98.2	525	10	US-09-919-0319-62
4	543	56.2	526	14	US-10-074-225A-3
5	325	33.6	67	12	US-10-424-599-166051
6	302	31.3	75	9	US-09-730-379B-1
7	280.5	29.0	101	14	US-10-074-225A-6
8	242.5	25.1	378	14	US-10-029-386-33892
9	221	22.9	292	9	US-09-864-761-37944
10	213	22.0	38	9	US-09-730-379B-9
11	213	22.0	38	9	US-09-730-379B-13
12	210	21.7	82	9	US-09-864-761-33313
13	199.5	20.7	130	16	US-10-389-566-2423
14	199.5	20.7	123	16	US-10-389-566-2424
15	199.5	20.7	265	16	US-10-389-566-2422

16	192.5	19.9	342	12	US-10-424-599-281824	Sequence 281824,
17	191.5	19.8	156	9	US-09-864-761-41679	Sequence 41679, A
18	185	19.2	26	9	US-09-809-657-2	Sequence 2, Appli
19	185	19.2	26	9	US-09-848-512A-2	Sequence 2, Appli
20	185	19.2	26	9	US-09-742-494-2	Sequence 4, Appli
21	185	19.2	26	14	US-10-153-312A-4	Sequence 37352, A
22	185	19.2	78	9	US-09-864-761-37352	Sequence 45941, A
23	182	18.8	253	12	US-10-282-122A-45941	Sequence 37327, A
24	179.5	18.6	87	9	US-09-864-761-37327	Sequence 34744, A
25	179.5	18.6	87	9	US-09-864-761-34744	Sequence 825, App
26	178.5	18.5	140	16	US-10-389-566-825	Sequence 69999, A
27	175	18.1	430	12	US-10-425-114-69999	Sequence 186228,
28	172	17.8	314	12	US-10-424-599-186228	Sequence 160290,
29	171.5	17.8	144	12	US-10-424-599-160290	Sequence 1014, Ap
30	166	17.2	153	14	US-10-083-357-1014	Sequence 130, App
31	164.5	17.0	695	10	US-09-890-688-130	Sequence 37882, A
32	164	17.0	43	9	US-09-864-761-37882	Sequence 31185, A
33	164	17.0	108	14	US-10-029-386-31185	Sequence 14106, A
34	164	17.0	574	14	US-10-156-761-14106	Sequence 33, Appli
35	162	16.8	476	15	US-10-360-849A-33	Sequence 76, Appli
36	160.5	16.6	366	12	US-10-406-686A-76	Sequence 122803,
37	158	16.4	139	16	US-10-437-963-122803	Sequence 33860, A
38	158	16.4	251	14	US-10-029-386-33860	Sequence 363, App
39	158	16.4	940	12	US-10-221-278-363	Sequence 198315,
40	158	16.4	940	15	US-10-231-172-363	Sequence 13228, A
41	157	16.3	2068	16	US-10-437-963-198315	Sequence 19, Appli
42	156.5	16.2	1130	14	US-10-032-585-7758	Sequence 4, Appli
43	155.5	16.1	421	14	US-10-156-761-13228	
44	155.5	16.1	865	9	US-09-957-995A-19	
45	155	16.0	29	9	US-09-730-379B-4	

## ALIGNMENTS

### RESULT 1

US-10-074-225A-5

; Sequence 5, Application US/10074225A

; Publication No. US20030082740A1

; GENERAL INFORMATION:

; APPLICANT: DONATE, Fernando

; APPLICANT: PLUNKETT, Marian L

; APPLICANT: HARRIS, Scott

; APPLICANT: MAZAR, Andrew P

; TITLE OF INVENTION: HISTIDINE PROLINE RICH GLYCOPROTEIN (HPRG) AS AN ANTI-ANGIOGENIC

; TITLE OF INVENTION: ANTI-TUMOR AGENT

; FILE REFERENCE: 38342-178463

; CURRENT APPLICATION NUMBER: US/10/074,225A

; PRIOR FILING DATE: 2002-02-14

; PRIOR FILING DATE: 2001-02-14

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 5

; LENGTH: 148

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-074-225A-5

Query Match 100.0%; Score 966; DB 14; Length 148;

Best Local Similarity 100.0%; Pred. No. 1.4e-69;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HPHKHSHEQHPGHGHPAHHEHDTROHPHGHGHPGHGHPGHGHPGHGHPCH 60

Db 1 HPHKHSHEQHPGHGHPAHHEHDTROHPHGHGHPGHGHPGHGHPGHGHPCH 60

Qy 61 DFQDYGCDPPNNQHCCHGHPFGHLRRRGKGPFPFHCQIGSVYRLPLKGEV 120

Db 61 DFQDYGCDPPNNQHCCHGHPFGHLRRRGKGPFPFHCQIGSVYRLPLKGEV 120

Qy 121 LPLPEANFPSPFLPHKHPKPDNOFPF 148





PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 37944  
LENGTH: 292  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC005414.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11  
US-09-864-761-37944  
Query Match 22.9%; Score 221; DB 9; Length 292;  
Best Local Similarity 44.1%; Pred. No. 5.5e-10;  
Matches 41; Conservative 3; Mismatches 31; Indels 18; Gaps 4;  
Qy 3 HKHSHEQPHGHHPHAPHEDHTHQPHGHHPGHHPGHHPGHHPCHD 61  
Db 26 HHHHHHDYDH-HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH- 83  
Qy 62 FDYGPDPHPHNOGHCCHGPPGCHRRRGP 94  
Db 84 -----PHHHHHHHH-----HRRHVP 101  
RESULT 10  
US-09-730-379B-9  
Sequence 9, Application US/09730379B  
Patent No. US20010041670A1  
GENERAL INFORMATION:  
APPLICANT: Simantov M.D., Ronit  
APPLICANT: Silverstein M.D., Roy L.  
TITLE OF INVENTION: THROMBOSPONDIN-BINDING REGION OF HISTIDINE-RICH  
FILE REFERENCE: 955-7P/CON  
CURRENT APPLICATION NUMBER: US/09730379B  
CURRENT FILING DATE: 2000-12-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37944  
LENGTH: 292  
TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION: PUBLICATION INFORMATION  
JOURNAL: Biochemistry  
VOLUME: 25  
ISSUE: 8  
PAGES: 2220-2225  
DATE: 1986  
DATABASE ACCESSION NUMBER: GenBank/P04196  
US-09-730-379B-9  
Query Match 22.9%; Score 213; DB 9; Length 38;  
Best Local Similarity 100.0%; Pred. No. 3.7e-10;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 97 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 134  
Db 1 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 38

FILE REFERENCE: 955-7P/CON  
CURRENT APPLICATION NUMBER: US/09730379B  
CURRENT FILING DATE: 2000-12-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 9  
LENGTH: 38  
TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION: PUBLICATION INFORMATION  
JOURNAL: Biochemistry  
VOLUME: 25  
ISSUE: 8  
PAGES: 2220-2225  
DATE: 1986  
DATABASE ACCESSION NUMBER: GenBank/P04196  
US-09-730-379B-9  
Query Match 22.0%; Score 213; DB 9; Length 38;  
Best Local Similarity 100.0%; Pred. No. 3.7e-10;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 97 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 134  
Db 1 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 38  
RESULT 11  
US-09-730-379B-13  
Sequence 13, Application US/09730379B  
Patent No. US20010041670A1  
GENERAL INFORMATION:  
APPLICANT: Simantov M.D., Ronit  
APPLICANT: Silverstein M.D., Roy L.  
TITLE OF INVENTION: THROMBOSPONDIN-BINDING REGION OF HISTIDINE-RICH  
FILE REFERENCE: 955-7P/CON  
CURRENT APPLICATION NUMBER: US/09730379B  
CURRENT FILING DATE: 2000-12-05  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 13  
LENGTH: 38  
TYPE: PRT  
ORGANISM: Homo sapiens  
PUBLICATION INFORMATION: PUBLICATION INFORMATION  
JOURNAL: Biochemistry  
VOLUME: 25  
ISSUE: 8  
PAGES: 2220-2225  
DATE: 1986  
DATABASE ACCESSION NUMBER: GenBank/P04196  
US-09-730-379B-13  
Query Match 22.0%; Score 213; DB 9; Length 38;  
Best Local Similarity 100.0%; Pred. No. 3.7e-10;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 97 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 134  
Db 1 GPRPFHCRQIGSVYRLPPLRKGEVLPPLPEANFSPFLP 38  
RESULT 12  
US-09-864-761-33313  
Sequence 33313, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Weisheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

```

; Publication NO. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2423
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-389-566-2423

Query Match      20.7%; Score 199.5; DB 16; Length 130;
Best Local Similarity 44.6%; Pred.No.1.3e-08;
Matches          45; Conservative 2; Mismatches 35; Indels 19; Gaps 8;

Qy    1 HPEKHSHSQHPHGHHPHAAHPPHEHTDRQDPH-GHHPH-GHHPH-PHGHP-----52
Db    9 HGHHGHHHQGHGGHGHGHHHHQHGGHGHGHHHHQGCHGCHHHHHQHGGHGHGHHGHHQQVH 68

Qy    53 -HGHHPHCHD-----FDYGFCDPPPNNQGCCHCGHGPFPQH 88
Db    69 HHGHH-HIHPQAIIYQTH-----QGHQHHDHRRGHGHGQH 103

RESULT_14
US-10-389-566-2424
; Sequence 2424, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2424
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-389-566-2424
```

[illegible]

69 HHGHH-HIHPOAILYOTH-----QGHQGHHDHHGHHGHQGH 103



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:23:24 ; Search time 68.9558 Seconds  
(without alignments)  
413.849 Million cell updates/sec

Title: US-10-074-225A-6

Perfect score: 697  
Sequence: 1 SVNIHRRPPHGHHPGPP.....HPPKGGHFDHGPCDPPSHK 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_29Jan04:\*
- 1: Geneseqp1980s:\*
  - 2: Geneseqp1990s:\*
  - 3: Geneseqp2000s:\*
  - 4: Geneseqp2001s:\*
  - 5: Geneseqp2002s:\*
  - 6: Geneseqp2003as:\*
  - 7: Geneseqp2003bs:\*
  - 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	697	100.0	526	5 ABB79805	Abb79805 Rabbit hi
2	314	45.1	525	5 ABB79804	Abb79804 Human pro
3	314	45.1	525	8 ADE76897	Adc76897 Human pro
4	280.5	40.2	148	5 ABB79807	Abb79807 Rabbit hi
5	280.5	40.2	148	5 ABB79806	Abb79806 Human his
6	264.5	37.9	491	4 ABB09949	Abb09949 Novel hum
7	259.5	37.2	296	7 ADD45442	Add45442 Rat Prote
8	259.5	37.2	296	7 ADE57105	Adc57105 Rat Prote
9	246	35.3	274	7 ADD47240	Add47240 Rat Prote
10	246	35.3	274	7 ADD48703	Add48703 Rat Prote
11	233.5	33.5	274	2 ABB04119	Aar04119 ORF2 of E
12	233.5	33.5	309	2 ABB04116	Aar04116 ORF1 of E
13	233.5	33.5	331	7 ADD45740	Add45740 Human pro
14	233.5	33.5	424	6 ABO52997	Abo52997 Human spl
15	231.5	33.2	192	6 AAO30246	Aao30246 Human sap
16	231.5	33.2	208	6 AAO30188	Aao30188 Human gp3
17	231	33.1	267	2 AAO30408	Aao30408 Human sec
18	229.5	32.9	124	2 AAW50192	Aaw50192 Amino aci
19	228.5	32.8	93	4 AAB71657	Aab71657 Human col
20	228.5	32.8	93	4 AAB71654	Aab71654 Human col
21	228.5	32.8	102	4 ABB68218	Abb68218 Drosophil
22	228	32.7	247	6 ABU11891	Abu11891 Human ABC
23	225	32.3	80	4 ABB61183	Abb61183 Drosophil
24	223	32.0	594	4 ABB61362	Abb61362 Drosophil
25	223	32.0	1130	5 ABB73921	Abb73921 Candida a

26	221	31.7	1378	5 ABB57176	Abb57176 Mouse isc
27	215.5	30.9	104	5 ABB78538	Abb78538 Ser-Pro-P
28	215	30.8	162	7 AAO30407	Aao30407 Human sec
29	210	30.1	487	2 AAR22380	Aar22380 Antigen m
30	209.5	30.1	205	5 ABR39650	Abr39650 A. thalia
31	208	29.8	865	3 AAB26153	Aab26153 Neospora
32	205	29.4	815	4 ABB70499	Abb70499 Drosophil
33	205	29.4	815	4 ABB70499	Abb70499 Drosophil
34	198.5	28.5	147	5 ABB43151	Abb43151 Human ova
35	198	28.4	164	7 ADD47283	Add47283 Rat Prote
36	196	28.1	82	2 AAW50193	Aaw50193 Amino aci
37	195.5	28.0	99	4 AAO02076	Aao02076 Human pol
38	195.5	28.0	99	4 AAO02076	Aao02076 Human pol
39	194.5	27.9	926	4 ABB65135	Abb65135 Drosophil
40	193	27.7	749	4 ABB62915	Abb62915 Drosophil
41	192.5	27.6	455	2 AAR12362	Aar12362 Octopus r
42	192.5	27.6	940	4 AAU28194	Aau28194 Novel hum
43	192	27.5	202	7 ADD48701	Add48701 Rat Prote
44	192	27.5	351	2 AAR24393	Aar24393 Sequence
45	192	27.5	464	6 ABO52933	Abo52933 Human spl

ALIGNMENTS

RESULT 1

ABB79805  
ID ABB79805 standard; protein; 526 AA.

XX ABB79805;  
AC  
XX  
XX  
DT 25-NOV-2002 (first entry)  
XX  
DE Rabbit histidine proline rich glycoprotein.  
XX  
KW Histidine proline rich glycoprotein; HPRG; rabbit; antiangiogenic;  
KW cystosatic; antiarteriosclerotic; antiinflammatory; antidiabetic;  
KW synaecological; antiarthritic; antiulcer; osteopathic; antitumor;  
KW ophthalmological; nontropic; neuroprotective; antiparkinsonian; chelator.  
XX  
OS Oryctolagus cuniculus.  
XX  
FH Key Location/Qualifiers  
FT Domain 251..296  
FT /note= "Proline-rich domain"  
FT Domain 321..421  
FT /note= "Histidine-proline-rich domain, region also  
FT specifically claimed in Claim 1"

WO200264621-A2.  
22-AUG-2002.

14-FEB-2002; 2002WO-US004336.

14-FEB-2001; 2001US-0268370P.

(ATTE-) ATTENUON LLC.

Donate F, Harris S, Plunkett ML, Mazar AP;

WPI; 2002-666989/71.

P-PSDB; ABBN84911.

New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for  
diagnosing or treating diseases associated with undesired cell migration,  
invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.

Claim 1; Page 12-13; 82pp; English.

The present sequence is the protein sequence of rabbit histidine proline  
rich glycoprotein (HPRG), a proteinaceous chelator that can be used to  
inhibit angiogenesis and treat cancer. Claimed anti-angiogenic

CC polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation in vitro or in vivo; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding molecule, or isolating or enriching cells expressing HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumours or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays.

XX Sequence 526 AA;

Query Match 100.0%; Score 697; DB 5; Length 526;  
Best Local Similarity 100.0%; Pred. No. 5.5e-51;  
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SVNIIHRPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPH 60  
Db 321 SVNIIHRPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPH 380

QY 61 GHPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPSHX 101  
Db 381 GHPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPPHGHPPSHX 421

RESULT 2  
ID ABB79804 standard; protein; 525 AA.  
XX ABB79804;  
XX ABB79804;  
XX 25-NOV-2002 (first entry)  
XX Human histidine proline rich glycoprotein.  
XX Histidine proline rich glycoprotein; HPRG; human; antiangiogenic; cytostatic; antiarteriosclerotic; antiinflammatory; antidiabetic; gynaecological; antiarthritic; antiulcer; osteopathic; antitumour; ophthalmological; nootropic; neuroprotective; antiparkinsonian; chelator.  
OS Homo sapiens.  
XX

Key Location/Qualifiers  
Domain 276..321  
Domain /note= "Proline-rich domain"  
Domain 350..497  
Domain /note= "Histidine-proline-rich domain, region also specifically claimed in Claim 1"

XX WO200264621-A2.  
XX 22-AUG-2002.  
XX 14-FEB-2002; 2002WO-US004336.  
XX 14-FEB-2001; 2001US-0268370P.  
XX (ATTE-) ATTENUON LLC.  
XX Donate F, Harris S, Plunkett ML, Mazar AP;  
XX WPI: 2002-666989/71.  
XX P-PSDB; ABN94910.  
XX New histidine-proline rich glycoprotein (HPRG) polypeptide, useful for diagnosing or treating diseases associated with undesired cell migration, invasion, proliferation, or angiogenesis, e.g. cancer, atherosclerosis.  
XX Claim 1; Page 11; 82pp; English.

CC The present sequence is the protein sequence of human histidine proline rich glycoprotein (HPRG), a proteinaceous metal chelator that can be used to inhibit angiogenesis and treat cancer. Claimed anti-angiogenic polypeptides or peptides comprise: the histidine-proline-rich (H/P) domain (see ABB79806) of human HPRG; an H/P domain of rabbit HPRG (see ABB79807); a variant of these that is capable of inhibiting angiogenesis, endothelial cell proliferation or endothelial tube formation in vitro or in vivo; or a pentapeptide having the generic sequence given in ABB79808. Also claimed are: chemically synthesised or recombinantly produced peptide multimers; a diagnostically or therapeutically labeled anti-angiogenic polypeptide, peptide or peptide multimer; a diagnostically useful HPRG-related composition, comprising the diagnostically labeled polypeptide, peptide or peptide multimer and a carrier; an antibody specific for an epitope of HPRG that is present in the H/P domain of human or rabbit HPRG, and which binds to HPRG or to any of the domains in a way which inhibits the anti-angiogenic activity of HPRG or the domain, or an antigen-binding fragment of the antibody; a method for inhibiting cell migration, cell invasion, cell proliferation or angiogenesis, or for inducing apoptosis; a method for treating a subject having a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis; a method of stimulating or inhibiting angiogenesis in a subject; a method of detecting the presence of HPRG or its cleavage product or its peptide in a biological sample; isolated nucleic acids encoding the polypeptide, peptide or peptide multimer; an expression vector; transformed or transfected cells; a method of providing to a cell, tissue or organ an angiogenesis-inhibitory amount of HPRG, an H/P domain of HPRG or its pentapeptide, or the peptide multimer; an affinity ligand useful for binding to, or isolating, an HPRG-binding molecule or cells expressing the binding molecule, comprising the polypeptide, peptide or peptide multimer, immobilised to a solid support or carrier; and a method of isolating HPRG-binding molecule, or isolating or enriching cells expressing HPRG-binding site or receptor, from a complex mixture. The compositions and methods are useful in diagnosing or treating a disease or condition associated with undesired cell migration, invasion, proliferation, or angiogenesis, such as cancer, atherosclerosis, diabetic retinopathy, inflammation, endometriosis, arthritis, peptic ulcers, or fractures. HPRG is especially useful in inhibiting the growth of primary tumours or metastases, and may also be used in treating neurodegenerative diseases like Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. The antibodies are stimulators of angiogenesis and are useful for promoting neovascularization in pertinent disease states, and in various immunoassays.

XX Sequence 525 AA;



Query Match	45.1%;	Score 314;	DB 5;	Length 525;
Best Local Similarity	41.8%;	Pred.No. 7.5e-19;		
Matches 64;	Conservative 2;	Mismatches 23;	Indels 64;	Gaps 8;
QY	7 RPP--PHG----	HEPHGPPPHGHHPHGPPP-	-----HGHP-PHGPPP-	-----40
Db	:			
	275 KPPFPKPGSRDHHHPKFP-	-----HEHGPPPPPPDERDHSHGPPLPQGPPLPMSCSQH	329	
QY	41	-----BHP-----	PHGPPPHGHP-	-----HGPPPHGHPHP 67
Db	330 ATFGTNGAQRSHNNNSD	LHPKHKSHSQHPGHHPHAAHHPHEDTHRQHPGHGHPGH 389		
QY	68 PPHGHPHPHPHPHPHPGHG	FGFHDHGFCDDPPSH 100		
Db	390 PPHGHPHPHPHPHPHPGHG	FGFHDHGFCDDPPSH 422		
RESULT 3				
ADE76897				
ID	ADE76897	standard; protein; 525 AA.		
XX	ADE76897;			
AC	XX			
DT	29-JAN-2004	(first entry)		
XX	XX	Human protein expressed in a liver disorder #18.		
DE	XX			
XX	XX	human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;		
KW	XX	tumour; liver; inflammatory disorder; immune response disorder;		
KW	XX	high-throughput screening; differential gene expression; gene therapy.		
XX	XX			
OS	XX	Homo sapiens.		
XX	XX			
PN	XX	US2003108871-A1.		
PD	XX	12-JUN-2003.		
PD	XX			
PF	XX	30-JUL-2001; 2001US-00919039.		
XX	XX			
PR	XX	28-JUL-2000; 2000US-0222113P.		
PA	XX	(KASE)/ KASER M R.		
XX	XX	Kaser MR;		
PI	XX			
XX	XX			
DR	XX	WPI; 2004-031227/03.		
DR	XX	N-PSDB; ADE76896.		
XX	XX			
PT	XX	Composition comprising several cDNAs that are differentially expressed in		
PT	XX	treated human C3A liver cell cultures, useful for treating liver		
PT	XX	disorders.		
XX	XX			
PS	XX	Claim 1; SEQ ID NO 62; 41pp; English.		
XX	XX			
CC	XX	The invention relates to a composition comprising several cDNAs that are		
CC	XX	differentially expressed in a liver disorder. The composition is useful		
CC	XX	for treating liver disorder such as hyperlipidaemia, hypertension, type		
CC	XX	II diabetes, tumours of the liver and disorders of the inflammatory and		
CC	XX	immune response. The composition is useful for a high-throughput method		
CC	XX	of screening several molecules or compounds to identify a ligand which		
CC	XX	specifically binds a cDNA. A protein encoded by the cDNA is useful for a		
CC	XX	high-throughput method for using a protein to screen several molecules or		
CC	XX	compounds to identify at least one ligand which specifically binds the		
CC	XX	protein which involves combining the protein encoded by the cDNA with		
CC	XX	several of molecules or compounds under conditions to allow specific		
CC	XX	binding, and detecting specific binding between the protein and a		
CC	XX	molecule or compound, therefore identifying a ligand which specifically		
CC	XX	binds the protein. The composition is useful for detecting and		
CC	XX	quantifying differential gene expression, can be used in gene therapy, to		
CC	XX	formulate prognosis and to design a treatment regimen and to monitor the		
CC	XX	efficacy of treatment. The present sequence represents the amino acid		
CC	XX	sequence of a protein encoded by a cDNA differentially expressed in a		



DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #9940.  
 XX  
 XX Human; Chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF  
 XX 31-MAR-2000; 2000US-00540217.  
 PR  
 XX 23-AUG-2000; 2000US-00649167.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 XX N-PSDB; AAS74136.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PT  
 PS Claim 20; SEQ ID NO 40308; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 XX Sequence 491 AA;  
 SQ  
 Query Match 37.9%; Score 264.5; DB 4; Length 491;  
 Best Local Similarity 41.8%; Pred. No. 18-14;  
 Matches 56; Conservative 2; Mismatches 35; Indels 41; Gaps 8;  
 QY 6 HRPP-----HGHPHG-----PPHGHHPH-GPPPHGHHPHPPPPHPPH----- 45  
 DB 136 HRPPPAQVTTSCGHQGHQHRPPPPAHHGHQSSPPAHHGHQHRPPAHHGHQHHQHS 195  
 QY 46 GPPPHGHHPHPPPHGHHPHGGP-----PHGHHPHGGPP-----HGHP-----PPHG 86  
 DB 196 SPPAHHGHQHRHQPCHGHQHRSPHSTAHQCHGHQSSPPAHHGHQHHQHHQHSPPYQ 255  
 QY 87 HGPHDHGCDPPSH 100  
 DB 256 HGHHQHRP--PPAH 267

RESULT 7  
 ADD45442  
 ID ADD45442 standard; protein; 296 AA.  
 XX  
 AC ADD45442;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Rat Protein L17318, SEQ ID NO 10875.  
 XX  
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX  
 OS Rattus norvegicus.  
 XX  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR  
 XX 01-NOV-2001; 2001US-0346382P.  
 PR  
 XX 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEHO) GEN HOSPITAL CORP.  
 PA  
 XX (FARB) BAYER AG.  
 XX  
 PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 XX WPI; 2003-268312/26.  
 DR  
 XX GENBANK; L17318.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 PT  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
 CC injury (CCI) and spared nerve injury (SNI)).  
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 296 AA;  
 SQ  
 Query Match 37.2%; Score 259.5; DB 7; Length 296;  
 Best Local Similarity 38.4%; Pred. No. 1.7e-14;  
 Matches 63; Conservative 1; Mismatches 27; Indels 73; Gaps 10;  
 QY 2 VNIHRPPPHG---HHPH-----GPPPHG-----HHPHGPPPHGHP-- 34



CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 274 AA;

Query Match 35.3%; Score 246; DB 7; Length 274;  
 Best Local Similarity 47.1%; Pred. No. 2.2e-13;  
 Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;

QY 8 PPHG--HHPHGPPPHG---HHPHGPPPHGHP---PHGPPP-----RHPPHGGPPHGHGHP- 53  
 DB 104 PPGGPGQQRPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPG 163  
 QY 54 --PHGPPPHG---HPPHGGPPHGHGHP---PHGPPPHGPPHGHGHP-----CDP 97  
 DB 164 QGPGGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPG 219

QY 98 P 98  
 DB 220 P 220

# RESULT 10

ADD48703  
 ID ADD48703 standard; protein; 274 AA.  
 AC ADD48703;  
 XX 29-JAN-2004 (first entry)  
 DT  
 DE Rat Protein AAA42064, SEQ ID NO 14412.  
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
 KW Chronic constriction injury; CCI; spared nerve injury; SNI; Chung.  
 XX Rattus norvegicus.  
 OS  
 PN WO2003016475-A2.  
 XX  
 XX 27-FEB-2003.  
 XX  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 XX WPI; 2003-268312/26.  
 DR GENBANK; AAA42064.  
 DR  
 XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 XX Claim 1; Page; 1017pp; English.  
 PS

XX The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 274 AA;

Query Match 35.3%; Score 246; DB 7; Length 274;  
 Best Local Similarity 47.1%; Pred. No. 2.2e-13;  
 Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;

QY 8 PPHG--HHPHGPPPHG---HHPHGPPPHGHP---PHGPPP-----RHPPHGGPPHGHGHP- 53  
 DB 104 PPGGPGQQRPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPG 163  
 QY 54 --PHGPPPHG---HPPHGGPPPHGHP---PHGPPPHGPPHGHGHP-----CDP 97  
 DB 164 QGPGGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPGPPPGPGQPG 219

QY 98 P 98  
 DB 220 P 220

# RESULT 11

AAR04119  
 ID AAR04119 standard; protein; 274 AA.  
 AC AAR04119;  
 XX  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 05-JUN-1990 (first entry)  
 XX  
 XX ORF2 of Enod2b genomic clone.  
 KW  
 KW ORF2; Enod2; soybean; root nodule; nodulin 75.  
 XX  
 OS Glycine max; (L) Merr. cv. Williams.  
 XX  
 XX EP349338-A.  
 XX  
 XX 03-JAN-1990.  
 XX  
 XX 01-JUL-1988; 88US-00214297.  
 XX  
 XX 01-JUL-1988; 88US-00214297.  
 XX  
 XX (LUBR ) LUBRIZOL GENETICS INC.



CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 331 AA;

Query Match 33.5%; Score 233.5; DB 7; Length 331;  
 Best Local Similarity 42.2%; Pred. No. 2.9e-12;  
 Matches 49; Conservative 4; Mismatches 38; Indels 25; Gaps 5;

QY 8 PPHGHHPHGGPPHGHHPHGGPPH-GHPHGGPPHPPHGGPPHGH-----PPHG 56  
 DB 101 PQGGNQPGPPPPPPKPGPPQGGNKGPPKPGPPKPGPPKPGPPKPGPPKPG 160

QY 57 PPHH-GHPHGGPP-----HGHPPHGGPPHGH-PPHGHHFHGHGCDPP 98  
 DB 161 PPQGGNQPGPPPPKPGPPKPGPPQGGNKGPPKPGPPKPGPPKPGPPKPG 216

RESULT 14

ID ABO52997 standard; protein; 424 AA.  
 AC ABO52997;  
 XX 09-OCT-2003 (first entry)  
 DT Human spliceosome associated protein (SAP) #24.  
 XX Human; SAP; spliceosome associated protein; ribonucleoprotein;  
 KW RNP complex; RNA affinity substrate; RNP assembly sequence;  
 KW spliceosomal complex; hnRNP complex; mRNA export complex;  
 KW mRNA localisation complex; RNA editing complex; intron complex;  
 KW H complex; telomerase complex; fragile X protein complex;  
 KW reverse transcriptase complex; gene splicing complex.  
 XX Homo sapiens.  
 OS US2003068803-A1.  
 PN 10-APR-2003.  
 PD 14-JAN-2002; 2002US-00047991.  
 XX 12-JAN-2001; 2001US-0261521P.  
 PR (REED/) REED R.  
 XX (ZHOU/) ZHOU Z.  
 PA Reed R, Zhou Z;  
 XX WPI; 2003-540885/51.  
 DR Isolating ribonucleoprotein complex, by contacting RNA affinity substrate  
 PT having ribonucleoprotein assembly sequence and affinity tag, with protein  
 PT mixture, subjecting complex formed to chromatography, affinity selection.  
 XX Claim 24; Page; 39pp; English.  
 PS The invention relates to forming (M1) an isolated ribonucleoprotein (RNP)  
 CC complex (C), involves contacting an RNA affinity substrate (S) comprising  
 CC an RNP assembly sequence (AS) and an affinity tag, with a protein mixture

CC to permit formation of (C) on AS, subjecting (C) to chromatographic  
 CC separation, and subjecting (C) to affinity selection, where the affinity  
 CC tag (e.g. bacteriophage MS2 coat protein in a fusion protein with E. coli  
 CC maltose binding protein) binds to an affinity matrix. Also included are  
 CC an isolated spliceosome preparation (isolated by (M1)), a RNA comprising  
 CC an RNP complex binding site and at least one phage coat protein  
 CC recognition site, a nucleic acid encoding the RNA, and treating (M2) a  
 CC subject having a disorder associated with abnormal RNP complexes (by  
 CC obtaining a sample of cells from a subject, purifying RNP complexes from  
 CC the cells of the subject by (M1), determining the presence in the  
 CC purified RNP complexes of one or more proteins, and normalising the  
 CC amount of RNPs in the subject. (M1) is useful for forming an isolated RNP  
 CC complex selected from a spliceosomal complex (selected from E, A, B and C  
 CC complex), an hnRNP complex, an mRNA export complex, an mRNA localisation  
 CC complex, an RNA editing complex, an intron complex, or an H complex. (M1)  
 CC is useful in a diagnostic assay for determining whether a subject has  
 CC abnormal RNP complexes. (M2) is useful for treating a subject having a  
 CC disorder associated with abnormal RNP complexes. (M1) is useful for  
 CC forming an isolated RNP complex such as a telomerase complex, a fragile X  
 CC protein complex, a reverse transcriptase complex or a gene splicing  
 CC complex. The present sequence represents a known human spliceosome  
 CC associated protein (SAP) isolated by the methods of the invention. Note:  
 CC The present sequence is not shown in the specification but was obtained  
 CC from Genbank or Swissprot using the information provided in table 1 of  
 CC the specification

XX Sequence 424 AA;

Query Match 33.5%; Score 233.5; DB 6; Length 424;  
 Best Local Similarity 46.0%; Pred. No. 3.5e-12;  
 Matches 57; Conservative 3; Mismatches 25; Indels 39; Gaps 11;

QY 9 PPHGH-PPHG-PP-----PPHG-PPHG-PPHG-PPHG-PPHG-PPHG-PPHG 46  
 DB 286 FGHGHSHPPPPPPGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGGPPHGG 345

QY 47 PPHGHPPHGGPP-----PPHG-PPHG-PPHG-PPHG-PPHG-PPHG-PPHG 94  
 DB 346 PPMGMPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 400

QY 95 CDPP 98  
 DB 401 LPPP 404

RESULT 15

ID AAO30246 standard; protein; 192 AA.  
 XX AAO30246;  
 AC AAO30246;  
 XX 03-SEP-2003 (first entry)  
 DT Human Saprip protein fragment #1.  
 DE Human; forensic analysis; chromosome marker; organelle-specific marker;  
 KW novel-related disorder; neurological disorder; gene therapy; nontropic;  
 KW neuroprotective; salivary proline-rich protein; Saprip.  
 XX Homo sapiens.  
 OS Synthetic.  
 OS WO2003046180-A2.  
 PN 05-JUN-2003.  
 PD 25-NOV-2002; 2002WO-EP013210.  
 XX 28-NOV-2001; 2001US-0334147P.  
 PR 14-DEC-2001; 2001US-0340465P.  
 PR 18-APR-2002; 2002US-0373947P.  
 XX (GEST ) GENSET SA.





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: September 23, 2004, 22:33:30 ; Search time 20.2811 Seconds  
(without alignments)  
479.034 Million cell updates/sec

Title: US-10-074-225A-6  
Perfect score: 697  
Sequence: 1 SVNTIHRPPHGHHPHGPPPP.....HPPHGHGFHDGCDPPPSHK 101

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	31.4	45.1	525	1 KGHUGH	histidine-rich gly
2	260.5	37.4	445	2 A60488	histidine-rich gly
3	259.5	37.2	295	2 B48013	proline-rich prote
4	250	35.9	128	2 D38355	basic proline-rich
5	248	35.6	278	2 B39066	proline-rich prote
6	247.5	35.5	206	1 P18T3	acidic proline-ric
7	246.5	35.4	227	2 C23149	proline-rich prote
8	246.5	35.4	240	2 A24264	proline-rich prote
9	246.5	35.4	317	2 A28996	proline-rich prote
10	239	34.3	240	2 B24264	proline-rich prote
11	236	33.9	330	2 S22140	nodulin Emod2 - Se
12	235.5	33.8	392	1 PIHUB6	salivary proline-r
13	235.5	33.8	432	2 T06782	extensin - soybean
14	235.5	33.8	580	2 T10863	extensin precursor
15	234.5	33.6	173	2 T51469	glycine/proline-ri
16	233.5	33.5	309	2 S08343	nodulin precursor
17	233.5	33.5	424	2 A54964	spliceosome-assoc
18	233.5	33.4	1571	2 T14155	zinc finger protei
19	232.5	32.9	251	1 PIHUPF	salivary proline-r
20	228.5	32.8	188	2 D29149	proline-rich prote
21	225.5	32.4	212	2 B36298	proline-rich prote
22	224	32.1	320	1 PIHUSD	salivary proline-r
23	222.5	31.9	220	2 A36298	proline-rich prote
24	221	31.7	309	2 S10889	proline-rich prote
25	221	31.7	1378	2 T30173	zinc finger protei
26	214.5	30.8	117	2 D40750	proline-rich prote
27	213.5	30.6	301	2 E29149	proline-rich prote
28	210.5	30.2	434	2 S11967	module-specific hy
29	210	30.1	237	2 S23737	proline-rich prote

30 210 30.1 440 1 S60755 rhodopsin - Allote  
31 210 30.1 620 2 S06733 hydroxyproline-ric  
32 209 30.0 97 2 S04554 nodulin - alfalfa  
33 208 29.8 223 2 A42817 proline-rich prote  
34 207 29.7 727 2 C84534 hypothetical prote  
35 206.5 29.6 112 2 S10101 modulation protein  
36 204.5 29.3 539 2 T28770 hypothetical prote  
37 202.5 29.1 170 2 A48013 proline-rich prote  
38 202 29.0 188 2 JH0481 basic proline-rich  
39 201 28.8 452 1 S14332 rhodopsin - northe  
40 199.5 28.6 172 2 B29149 proline-rich prote  
41 198.5 28.5 1176 2 T49482 hypothetical prote  
42 198 28.4 164 2 A30496 proline-rich prote  
43 197.5 28.3 260 2 S22373 proline-rich prote  
44 197.5 28.3 592 2 T32402 hypothetical prote  
45 197.5 28.3 1002 2 T43236 carboxypeptidase C

ALIGNMENTS

RESULT 1  
KGHUGH  
histidine-rich glycoprotein precursor - human  
N:Alternate names: HRG  
C:Species: Homo sapiens (man)  
C>Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 16-Jun-2000  
C:Accession: A01287; S29669  
R:Koide, T.; Foster, D.; Yoshitake, S.; Davie, E.W.  
Biochemistry 25, 2220-2225, 1986  
A:Title: Amino acid sequence of human histidine-rich glycoprotein derived from the nuclei of human placental cells  
A:Reference number: A01287; MUID:86216149; PMID:3011081  
A:Accession: A01287  
A:Molecule type: mRNA  
A:Residues: 1-525 <KOI>  
A:Cross-references: GB:AB005803; NID:g2280513; PIDN:BAA21613.1; PID:g2280514  
R:Hennis, B.; Havelaar, A.; Klufft, C.  
submitted to the EMBL Data Library, October 1991  
A:Description: PCR detection of a dinucleotide repeat in the human histidine-rich glycoprotein  
A:Reference number: S29669  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 214-247 <HEN>  
A:Cross-references: EMBL:Z17218; NID:g32453; PIDN:CAA78925.1; PID:g32454  
C:Comment: Although its physiological function is not yet known, HRG does bind heme, dyes, and the lysine-binding site of plasminogen. On the basis of its homology with HRW, a blood coagulation cascade.  
C:Comment: The amino half of this protein is homologous to the first two cystatin-like proteins, which could not have inhibitory activity.  
C:Comment: In addition to having a high histidine and proline content, this protein has a high 'rich' region.  
A:Gene: GDB:HRG  
A:Cross-references: GDB:i20055; OMIM:142640  
A:Map position: 3q27-3q27  
A:Superfamily: histidine-rich glycoprotein; cystatin homology  
C:Keywords: duplication; glycoprotein; heparin binding; tandem repeat  
F1-18/Domain: signal sequence #status predicted <SIG>  
F19-525/Product: histidine-rich glycoprotein #status predicted <MAT>  
F19-131/Domain: cystatin homology <CY1>  
F140-246/Domain: cystatin homology <CY2>  
F1276-321/Region: proline-rich  
F1351-497/Region: histidine-rich  
F1351-497/Region: proline-rich  
F163125,344,345/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F178-89,105-126,218-241/Disulfide bonds: #status predicted

Query Match 45.1%; Score 314; DB 1; Length 525;  
Best Local Similarity 41.8%; Pred. No. 1.1e-16;  
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;  
QY 7 RPP--PHG----HPPHGHPPHGHHPHGPPPP-----HGHP-PHGPPPP----- 40

Db 275 KPFFKPGSRDHHHPKP-----HEGPPPPPPDERHSHGFPFLPQGPFPPLPMSCSSCOH 329  
 Qy 41 -----RHP-----PHGPPPHGPP-----HGPPPHGPPHGP 67  
 Db 330 ATFGTGAQRHSHNNSSDLPHPHKHSHEQPHGHPHFAHHPHEDTHRQHHPGHHPGH 389  
 Qy 68 PHGHPHGPHPGHPHGHGHDGCPDPSPH 100  
 Db 390 HPGHHPHGHHPHGHHPCHDFODYDGPDPSPH 422

RESULT 2  
 A: Molecule type: protein  
 A: Residues: 1-25; 26-52; 57-75; 82-88; 95-119; 146-173; 175-206; 210-309; 313-445 <SOR>  
 A: Note: 355-Gln and 368-Tyr were also found  
 R: Halkier, T.; Andersen, H.; Vestergaard, H.; Magnusson, S.  
 Biochem. Biophys. Res. Commun. 200, 78-82, 1994  
 A: Title: Bovine histidine-rich glycoprotein is a substrate for bovine plasma factor XIII  
 A: Reference number: JC2196; MUID: 94220160; PMID: 7909439  
 A: Accession: JC2196  
 A: Molecule type: protein  
 A: Residues: 1-23; 35-54; 'VK', 57-101, 'R', 'TVGEYTEG', 116, 'N', 118, 'R', 120-136; 137-145; 150-20  
 R: Vestergaard, A.B.; Andersen, H.F.; Magnusson, S.; Halkier, T.  
 Thromb. Res. 60, 385-396, 1990  
 A: Title: Histidine-rich glycoprotein inhibits contact activation of blood coagulation.  
 A: Reference number: A60488; MUID: 91196010; PMID: 2084959  
 A: Accession: A60488  
 A: Molecule type: protein  
 A: Residues: 1-6; 'X', 8-15 <VES>  
 C: Comment: This protein is a single-chained plasma protein which participates in transglu  
 C: Superfamily: histidine-rich glycoprotein; cystatin homology  
 C: Keywords: glycoprotein; plasma  
 F: 1-13/Domain: cystatin homology <CY1>  
 F: 122-207/Domain: cystatin homology (fragments) <CY2>  
 F: 7-424, 60-71, 87-108, 165-346, 180-203, 258-288/Disulfide bonds: #status experimental  
 F: 74, 107, 164, 266/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 37.4%; Score 260.5; DB 2; Length 445;  
 Best Local Similarity 42.4%; Pred. No. 8.7e-13;  
 Matches 53; Conservative 4; Mismatches 33; Indels 35; Gaps 6;  
 Qy 11 HGHHPHGPSP-----HGHPHG-----PPPHGHPHG-----PPP--- 40  
 Db 229 HEHSPAGPFPKPSGSKDHG-HPHESYNFCPPPLEKHNSDPSPPFQARAPLPFPPLGLR 287  
 Qy 41 -RHPPHGPFPHPHPH---GPPPHGHPHPHPHPHPHPHPHPHPHPHPHPHPHPHPHP 96  
 Db 288 CPHPFPPTKGNRPPPHDSDPHNHPHHPHHPHPHPHPHPHPHPHPHPHPHPHPHPHP 347  
 Qy 97 PPSHX 101  
 Db 348 PPHR 352

RESULT 3  
 B48013  
 A: Title: proline-rich proteoglycan 2 precursor, parotid - rat  
 C: Species: Rattus norvegicus (Norway rat)  
 C: Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 20-Aug-1999  
 C: Accession: B48013

R: Castle, A.M.; Castle, J.D.  
 J. Biol. Chem. 268, 20490-20496, 1993  
 A: Title: Novel secretory proline-rich proteoglycans from rat parotid. Cloning and charac  
 A: Reference number: A48013; MUID: 9338626; PMID: 8376404  
 A: Accession: B48013  
 A: Status: preliminary  
 A: Molecule type: mRNA  
 A: Residues: 1-295 <CAS>  
 A: Cross-references: GB: L17318; NID: g310199; PIDN: AAA03074.1; PID: g310200  
 C: Superfamily: proline-rich protein  
 C: Keywords: extracellular protein; glycoprotein; tandem repeat

Query Match 37.2%; Score 259.5; DB 2; Length 295;  
 Best Local Similarity 38.4%; Pred. No. 7.3e-13;  
 Matches 63; Conservative 1; Mismatches 27; Indels 73; Gaps 10;  
 Qy 2 VNTHRPPPHG---HHPH-----GPPPHG-----HHHPHPHPHPHP--- 34  
 Db 81 VNRPERPFGGNNHHPHPPAAGPQRPPQPGSPQGGPQQRPPQPGPQGGPQQR 140  
 Qy 35 -----PHGPPP-----RHPPHGPHPHPH-----PHGPPPHGHP--- 71  
 Db 141 PPQPGSPGPPPGPQQRPPQPGPQGGPQQRPPQPGSPQGGPQQRPPQPGPQGGP 200  
 Qy 72 HP-----PHGP-----PPHHPHGHGHDHGHGCDPP 98  
 Db 201 GPQRPPQPGSPGPPPGPQQRPPQPGPQGGPQGGPQGGPQGGPQGGPQGGP 238

RESULT 4  
 B38355  
 A: Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A: Reference number: A38355; MUID: 91190884; PMID: 1849422  
 A: Accession: D38355  
 A: Status: preliminary  
 A: Molecule type: protein  
 A: Residues: 1-128 <KAU>  
 C: Superfamily: proline-rich protein

Query Match 35.9%; Score 250; DB 2; Length 128;  
 Best Local Similarity 45.0%; Pred. No. 1.8e-12;  
 Matches 50; Conservative 5; Mismatches 36; Indels 20; Gaps 5;  
 Qy 8 PPHGHP 60  
 Db 10 PPGGNGPQGGPP 69  
 Qy 61 GHPPHPPPPH-GHPHPGPPPHGH-----PPHGHGFHDHGFCDPPS 99  
 Db 70 GGRPQGPQGGKNGKPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 119

RESULT 5  
 B39066  
 A: Title: proline-rich protein 15 - rat  
 C: Species: Rattus norvegicus (Norway rat)  
 C: Date: 20-Mar-1992 #sequence\_revision 10-Apr-1992 #text\_change 03-May-1996  
 C: Accession: B39066  
 R: Lin, H.H.; Ann, D.K.  
 Genomics 10, 102-113, 1991  
 A: Title: Molecular characterization of rat multigene family encoding proline-rich protei  
 A: Reference number: A39066; MUID: 91257817; PMID: 2045095  
 A: Accession: B39066  
 A: Status: preliminary  
 A: Molecule type: DNA  
 A: Residues: 1-278 <LIN>  
 A: Cross-references: GB: M36414

C:Superfamily: proline-rich protein

Query Match 35.6%; Score 248; DB 2; Length 278;  
 Best Local Similarity 47.2%; Pred. No. 4.9e-12;  
 Matches 58; Conservative 1; Mismatches 32; Indels 32; Gaps 9;  
 QY 8 PPPHG-----HHPHGPPPHG-----HHPHGPPPHG-----HHPHGPPPHG 51  
 Db 52 PPKGPPQPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 111  
 QY 52 HP-----PHGPPPHGHP-----PHGPPPHG-----HHPHG-----HGFHDHGPC 95  
 Db 112 GPQPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 171  
 QY 96 DPP 98  
 Db 172 GPP 174

RESULT 6  
 PIRTS  
 acidic proline-rich protein precursor - rat  
 N:Alternate names: PRP  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 22-Jun-1999  
 C:Accession: A03296  
 R:Ziemer, M.A.; Swain, W.F.; Rutter, W.J.; Clements, S.; Ann, D.K.; Carlson, D.M.  
 J. Biol. Chem. 259, 10475-10480, 1984  
 A:Title: Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homolog  
 A:Reference number: A03296; MUID:84289443; PMID:6547951  
 A:Accession: A03296  
 A:Molecule type: mRNA  
 A:Residues: 1-206 <21E>  
 A:Cross-references: GB:K02247; NID:Q206395; PIDN:AAA41949.1; PID:G206396  
 C:Comment: This protein contains six 18- to 19-residue repeats.  
 C:Comment: This protein may protect teeth by binding to tannins.  
 C:Superfamily: proline-rich protein  
 C:Keywords: duplication; parotid gland; saliva; tandem repeat  
 F:1-13/Domain: signal sequence #status predicted <SIG>  
 F:14-206/Product: acidic proline-rich protein #status predicted <MAT>  
 F:80-189/Region: 18-residue repeats

Query Match 35.5%; Score 247.5; DB 1; Length 206;  
 Best Local Similarity 48.2%; Pred. No. 4.1e-12;  
 Matches 53; Conservative 4; Mismatches 30; Indels 23; Gaps 7;  
 QY 7 RPPPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 58  
 Db 79 RPPPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGHHPHGH 135  
 QY 59 PHGHP-----PHGPPPHGPPPHGHP-----PHGPPPHGPPPHGPPPHGPP 97  
 Db 136 PQGPPQPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 184

RESULT 7  
 C29149  
 proline-rich protein - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 08-Mar-1989 #sequence\_revision 08-Mar-1989 #text\_change 20-Aug-1999  
 C:Accession: C29149  
 R:Clements, S.; Mehansho, H.; Carlson, D.M.  
 J. Biol. Chem. 260, 13471-13477, 1985  
 A:Title: Novel multigene families encoding highly repetitive peptide sequences. Sequence  
 A:Contents: Clone pUMP40  
 A:Accession: C29149  
 A:Molecule type: mRNA  
 A:Residues: 1-227 <CUE>  
 A:Cross-references: GB:M11902; NID:G200544; PIDN:AAA40003.1; PID:G200545  
 C:Superfamily: proline-rich protein

Query Match 35.4%; Score 246.5; DB 2; Length 227;

Best Local Similarity 43.6%; Pred. No. 5.3e-12;  
 Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;  
 QY 7 RPPPHGHHP-----HGPDPHG-----HHPHGPPPHG-----HHPHGPPPHG 45  
 Db 33 RPPSGPPPPPPVNGSQGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 92  
 QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPPPHG 86  
 Db 93 GPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 152  
 QY 87 -HGFHDHGPCDPP 98  
 Db 153 PAGQPPRPPGPP 165

RESULT 8  
 A24264  
 proline-rich protein MP2 - mouse (fragment)  
 C:Species: Mus musculus (house mouse)  
 C:Date: 09-Sep-1987 #sequence\_revision 09-Sep-1987 #text\_change 03-May-1996  
 C:Accession: A24264  
 R:Ann, D.K.; Carlson, D.M.  
 J. Biol. Chem. 260, 15863-15872, 1985  
 A:Title: The structure and organization of a proline-rich protein gene of a mouse multi;  
 A:Reference number: A2508; MUID:86059475; PMID:2999141  
 A:Accession: A24264  
 A:Molecule type: DNA  
 A:Residues: 1-240 <ANN>  
 C:Superfamily: proline-rich protein

Query Match 35.4%; Score 246.5; DB 2; Length 240;  
 Best Local Similarity 43.6%; Pred. No. 5.6e-12;  
 Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;  
 QY 7 RPPPHGHHP-----HGPDPHG-----HHPHGPPPHG-----HHPHGPPPHG 45  
 Db 12 RPPSGPPPPPPVNGSQGPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 71  
 QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPPPHG 86  
 Db 72 GPPPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 131  
 QY 87 -HGFHDHGPCDPP 98  
 Db 132 PAGQPPRPPGPP 144

RESULT 9  
 A28996  
 proline-rich protein M14 precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 20-Aug-1999  
 C:Accession: A28996  
 R:Ann, D.K.; Smith, K.; Carlson, D.M.  
 J. Biol. Chem. 263, 10887-10893, 1988  
 A:Title: Molecular evolution of the mouse proline-rich protein multigene family. Insert;  
 A:Reference number: A28996; MUID:88273214; PMID:2839509  
 A:Accession: A28996  
 A:Molecule type: DNA  
 A:Residues: 1-317 <ANN>  
 A:Cross-references: GB:M23236; GB:J03891; NID:G200535; PIDN:AAA53048.1; PID:G567232  
 C:Genetics:

A:Introns: 22/1  
 C:Superfamily: proline-rich protein  
 C:Keywords: saliva  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-317/Product: proline-rich protein M14 #status predicted <MAT>

Query Match 35.4%; Score 246.5; DB 2; Length 317;  
 Best Local Similarity 43.6%; Pred. No. 7.1e-12;  
 Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;

```

QY 7 RPPPHGHP-----HGPPPHG-----HHPHGPPPHG-----HPPHGPP-----RHPPH 45
Db 33 RPPSGSQPRPPVNGSQGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPP 92
QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHG 86
Db 93 GPPPPGPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPP 152
QY 87 -HGPHDHPGCDPP 98
Db 153 PGQPQPPPPGPP 165

RESULT 10
B24264
Proline-rich protein MP3 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 09-Sep-1987 #sequence_revision 09-Sep-1987 #text_change 03-May-1996
C:Accession: B24264
R:Ann, D.K.; Carlson, D.M.
J. Biol. Chem. 260, 15863-15872, 1985
A:Title: The structure and organization of a proline-rich protein gene of a mouse multi
A:Reference number: A92508; MUID:86059475; PMID:2999141
A:Accession: B24264
A:Molecule type: DNA
A:Residues: 1-240 <ANN>
C:Superfamily: proline-rich protein

Query Match 34.3%; Score 239; DB 2; Length 240;
Best Local Similarity 42.6%; Pred. No. 2e-11;
Matches 58; Conservative 0; Mismatches 32; Indels 46; Gaps 10;

QY 7 RPPPHGHP-----HGPPPHG-----HPPHGPPPHG-----HPPHGPP-----RHPPH 45
Db 12 RPPSGSQPRPPVNGSQGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPPPPGPPGPP 71
QY 46 GPPPHG-----HPPHGPPPHG-----HPPHGPPPHG-----HPPHGPP-----PHGHPHG----- 86
Db 72 GPPPPGPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPP 131
QY 87 ---HGPHDHPGCDPP 98
Db 132 PAGPHLRPTQGP--PP 145

RESULT 11
S22140
nodulin Enod2 - Sesbania rostrata
C:Species: Sesbania rostrata
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999
C:Accession: S22140
R:de Bruijn, F.J.
submitted to the EMBL Data Library, November 1991
A:Reference number: S22140
A:Accession: S22140
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <DEB>
A:Cross-references: EMBL:X63339; NID:G21372; PIDN:CAA44939.1; PID:G21373
C:Superfamily: proline-rich protein 3

Query Match 33.9%; Score 236; DB 2; Length 330;
Best Local Similarity 41.1%; Pred. No. 4.4e-11;
Matches 44; Conservative 10; Mismatches 41; Indels 12; Gaps 3;

QY 7 RPPPHGHPPPPHGHPHGPPHG-----PPPHGHPHGPPPHR-----PPHGPPPHGHPHGPPHG 56
Db 151 KPPPEYQPPHKKPPPPYQPSKPPPPYPPYKPPHKKPPSPYKPPYKPPHKKPPYK 210
QY 57 RPPPHGHPPPPHGHPHGPPPHGHPHGHPHGHPHGHPHGHPHGHPHGHPHGHPHGHPHGHPHGHPHG 101
Db 211 KPPHKKPPHKKPPYKPPYKPPHKKPPHKKPPHKKPPHKKPPHKKPPHKKPPHKKPPHKKPPHKK 257

```

## RESULT 12

```

PIHUB6
salivary proline-rich phosphoprotein precursor PRB1 (large allele) [validated] - human
N:Contains: peptide IB-1; peptide P-E (peptide IB-9); peptide P-F; peptide P-H
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1986 #sequence_revision 12-Apr-1996 #text_change 08-Dec-2000
C:Accession: B40750; C40750; A40750; C25372; S02128; S02127; A03293; A90502; A91974; A05
A:Zelen, E.A.; Latreille, P.; Niece, R.L.
Am. J. Hum. Genet. 53, 264-278, 1993
A:Title: PRB1 gene variants coding for length and null polymorphisms among human saliv
A:Reference number: A40750; MUID:93304421; PMID:8317492
A:Accession: B40750
A:Molecule type: DNA
A:Residues: 35-392 <AZE>
A:Cross-references: GB:S62941
A:Experimental source: subject C.J. (large allele)
A:Accession: C40750
A:Molecule type: DNA
A:Residues: 35-127, 'R', 129-148, 'R', 150-151, 153-187, 'K', 189-272, 'S', 274-336, 'S', 338-392 <
A:Cross-references: GB:S62929
A:Experimental source: subject M.V.O. (large allele)
A:Accession: A40750
A:Molecule type: DNA
A:Residues: 35-183, 245-270, 'Q', 272-392 <AZ3>
A:Cross-references: GB:S62928
A:Experimental source: subject C.J. (medium allele)
A:Note: authors translated the codon CAA for residue 272 as Arg
J. Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A:Title: Differential RNA splicing and post-translational cleavages in the human saliv
A:Reference number: A92492; MUID:85289325; PMID:2993301
A:Accession: C25372
A:Molecule type: mRNA
A:Residues: 1-183, 245-392 <MAE>
A:Cross-references: GB:X03204; NID:G190485; PIDN:AAA60185.1; PID:G190486
A:Note: alternatively splice forms lacking portions of the repeat region were also found
R. Lyons, K.M.; Stein, J.H.; Smithies, O.
Genetics 120, 267-278, 1988
A:Title: Length polymorphisms in human proline-rich protein genes generated by intrageni
A:Reference number: S02127; MUID:89121440; PMID:2851479
A:Accession: S02128
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-127, 250-273, 'R', 275-277, 'R', 279-336, 'S', 338-392 <LYO>
A:Cross-references: EMBL:X07517
A:Accession: S02127
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 35-183, 245-392 <LY2>
A:Cross-references: EMBL:X07516
R. Kauffman, D.; Hofmann, T.; Bennick, A.; Keller, P.
Biochemistry 25, 2387-2392, 1986
A:Title: Basic proline-rich proteins from human parotid saliva: complete covalent struct
A:Reference number: A90502; MUID:86243355; PMID:3521730
A:Accession: A03293
A:Molecule type: protein
A:Residues: 17-36, 'AP', 41-51, 52-148, 'R', 150-152 <KA2>
A:Note: among nine basic proline-rich peptides isolated from the saliva, this peptide is
A:Accession: A90502
A:Molecule type: protein
A:Residues: 275-336, 'S', 338-392 <KAU>
R. Saitoh, E.; Isemura, S.; Sanada, K.
J. Biochem. 94, 1991-1999, 1983
A:Title: Further fractionation of basic proline-rich peptides from human parotid saliva
A:Reference number: A91974; MUID:84161824; PMID:6571974
A:Contents: P-H
A:Accession: A91974
A:Molecule type: protein
A:Residues: 5', 338-392 <SAI>
R. Azen, E.; Lyons, K.M.; McGonigal, T.; Barrett, N.L.; Clements, L.S.; Maeda, N.; Vanin,
Proc. Natl. Acad. Sci. U.S.A. 81, 5561-5565, 1984
A:Reference number: A94005; MUID:84298176; PMID:6089212

```

Query Match 33.8%; Score 235.5; DB 2; Length 432;  
 Best Local Similarity 41.7%; Pred. No. 5.9e-11;  
 Matches 58; Conservative 2; Mismatches 32; Indels 47; Gaps 14;

QY 6 HRPDPGHHP-----HGPPHGHHP-----HGPPP-HGHPP---HGPPPRHPHPGP 47  
 DB 49 HSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPP---PKHSP 106  
 QY 48 PP-----HGHP--HGPPP---HGHP--HGPPP---HGHP--HGPPP---HGHP--HGPPP 85  
 DB 107 PPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPP 166  
 QY 86 GHG-----FHDHGCDPPSH 100  
 DB 167 KHSPPPPYYHSP-PPPKH 184

RESULT 14  
 T10863  
 extensin precursor - kidney bean  
 N;Alternate names: cell wall protein; hydroxyproline-rich glycoprotein  
 C;Species: Phaseolus vulgaris (kidney bean)  
 C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 17-Nov-2000  
 R;Accession: T10863; B29356  
 R;Wycoff, K.L.; Powell, P.A.; Gonzales, R.A.; Corbin, D.R.; Lamb, C.; Dixon, R.A.  
 plant Physiol. 109, 41-52, 1995  
 A;Title: Stress activation of a bean hydroxyproline-rich glycoprotein promoter  
 A;Reference number: Z17192; MUID:96061709; PMID:7480331  
 A;Accession: T10863  
 A;Status: preliminary; translated from GB/EMBL/DBDJ  
 A;Molecule type: DNA  
 A;Residues: 1-590 <WYC>  
 A;Cross-references: EMBL:U18791; NID:g727263; PIDN:AAA87902.1; PID:g727264  
 A;Experimental source: strain Tendergreen  
 R;Corbin, D.R.; Sauer, N.; Lamb, C.J.  
 Mol. Cell. Biol. 7, 4337-4344, 1987  
 A;Title: Differential regulation of a hydroxyproline-rich glycoprotein gene fan  
 A;Reference number: A29356; MUID:88142825; PMID:3437892  
 A;Accession: B29356  
 A;Molecule type: mRNA  
 A;Residues: 'N', 1-229 <COR>  
 A;Cross-references: EMBL:M18094; NID:g169346; PIDN:AAA33764.1; PID:g169347  
 A;Experimental source: cv. Kievitboon Koekoek  
 C;Superfamily: hydroxyproline-rich glycoprotein  
 C;Keywords: glycoprotein; hydroxyproline  
 F;1-29/Domain: signal sequence #status predicted <SIG>  
 F;30-590/Product: extensin #status predicted <MAT>

Query Match 33.8%; Score 235.5; DB 2; Length 580;  
 Best Local Similarity 41.7%; Pred. No. 7.6e-11;  
 Matches 58; Conservative 2; Mismatches 32; Indels 47; Gaps 14;

QY 6 HRPDPGHHP-----HGPPHGHHP-----HGPPP-HGHPP---HGPPPRHPHPGP 47  
 DB 61 HSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPP---PKHSP 118  
 QY 48 PP-----HGHP--HGPPP---HGHP--HGPPP---HGHP--HGPPP---HGHP--HGPPP 85  
 DB 119 PPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPPPKHSPPPPYYHSPPP 178  
 QY 86 GHG-----FHDHGCDPPSH 100  
 DB 179 KHSPPPPYYHSP-PPPKH 196

RESULT 15  
 T51469  
 glycine/proline-rich protein - Arabidopsis thaliana  
 N;Alternate names: protein K10A8.130  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 18-Aug-2000  
 C;Accession: T51469

submitted to the Protein Sequence Database, August 2000

A;Reference number: Z25394  
 A;Accession: T51469  
 A;Status: Preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-173 <SAT>  
 A;Cross-references: EMBL:AL391151  
 A;Experimental source: cultivar Columbia; BAC clone K10A8  
 C;Genetics:  
 A;Map position: 5  
 A;Introns: 97/1  
 A;Note: K10A8\_130

Query Match 33.6%; Score 234.5; DB 2; Length 173;  
 Best Local Similarity 51.7%; Pred. No. 3.3e-11;  
 Matches 45; Conservative 7; Mismatches 22; Indels 13; Gaps 5;  
 Qy 9 PPHGHHPHGPPPHGHHPHGPPPHGHP-PHGPPRRPHGPPPHGHP-PHGHPPHG 66  
 Db 26 PPHGH-----GYGHHGHG---YGSSYYPPPP--PPHGYPPVAYPPHGYPPAGYPPAG 74  
 Qy 67 PPHGHPHPGPPPHGHPHPGHPGPHDGH 93  
 Db 75 YPPAGYFAGYPSHGYPSPSHGHHG 101

Search completed: September 23, 2004, 22:55:57  
 Job time : 20.2811 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 21:27:50 ; Search time 11.7631 Seconds  
(without alignments)  
447.085 Million cell updates/sec

Title: US-10-074-225A-6  
Perfect score: 697  
Sequence: 1 SVNIHRPPHGHGHPGPP.....HPPHGHGHDGCDPPSHK 101

Scoring table: BLOSUM62  
Gapop'10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	697	100.0	526	1 HRG RABIT	Q28640 cryotolagus
2	314	45.1	525	1 HRG HUMAN	P04196 homo sapien
3	260	37.3	396	1 HRG BOVIN	P33433 bos taurus
4	247.5	35.5	206	1 PRP3 RAT	P04474 rattus norv
5	246.5	35.4	261	1 PRP2 MOUSE	P05142 mus musculu
6	246.5	35.4	296	1 PMP3 MOUSE	P05143 mus musculu
7	233.5	33.5	309	1 NO75 SOYBN	P08297 glycine max
8	233.5	33.5	331	1 PRP1 HUMAN	P04280 homo sapien
9	233.5	33.5	424	1 S3B4 HUMAN	Q15427 homo sapien
10	229.5	32.9	251	1 PRP2 HUMAN	P02812 homo sapien
11	228	32.7	247	1 PRB4 HUMAN	P10163 homo sapien
12	224	32.1	276	1 PRPL HUMAN	P10162 homo sapien
13	221.5	31.8	234	1 PRPM HUMAN	P10161 homo sapien
14	210.5	30.2	434	1 NO75 LUPLU	Q06841 lupinus lut
15	210	30.1	620	1 EXTN TOBAC	P13983 nicotiana t
16	209	30.0	97	1 NO75 MEDSA	P11728 medicago sa
17	206.5	29.6	112	1 NO75 PRA	P16329 pisum sativ
18	205	29.4	815	1 PYGO DROME	Q9V988 drosophila
19	201	28.8	439	1 OPSD LOLSU	Q17094 loligo subu
20	201	28.8	452	1 OPSD LOLFO	P24603 loligo forb
21	199.5	28.6	172	1 PRP2 RAT	P10164 rattus norv
22	197.5	28.3	475	1 S3A2 MOUSE	Q62203 mus musculu
23	197.5	28.3	1002	1 CBPV SCHPO	O13849 schizosacch
24	192.5	27.6	455	1 OPSD OCTDO	P09241 octopus dof
25	192	27.5	331	1 HRPX PLALO	P04929 plasmodium
26	192	27.5	454	1 S3A2 HUMAN	Q15428 homo sapien
27	191	27.4	174	1 PRPP HUMAN	P81489 homo sapien
28	190	27.3	166	1 PRP3 HUMAN	P02810 loligo forb
29	188.5	27.0	96	1 PRP5 HUMAN	P04281 homo sapien
30	187	26.8	215	1 STRD MOUSE	Q60924 mus musculu
31	184.5	26.5	373	1 EXTL ARATH	Q38913 arabidopsis
32	184.5	26.5	464	1 OPSD SEPOF	O16005 sepiia offic
33	183	26.3	2716	1 OSA_DROME	Q81894 drosophila

## ALIGNMENTS

### RESULT 1

ID	HRG RABIT	STANDARD;	PRT;	526 AA.
AC	Q28640;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Histidine-rich glycoprotein precursor (Histidine-proline rich glycoprotein) (HPRG) (Fragment).			
DE	Glycoprotein (HPRG) (Fragment).			
GN	HRG.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_TaxID=9986;			
RN	[1]_TaxID=9986;			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 9-23; 301-313 AND 422-429.			
RC	TISSUE=Serum;			
RX	MEDLINE=96229917; PubMed=8639676;			
RA	Borza D.-B., Tatum F.M., Morgan W.T.;			
RA	"Domain structure and conformation of histidine-proline-rich glycoprotein.";			
RT	Glycoprotein.";			
RL	Biochemistry 35:1925-1934(1996).			
CC	FUNCTION: The physiological function is not yet known. It binds heme, dyes and divalent metal ions. It can inhibit rosette formation and is known to interact with heparin, thrombospondin, and the lysine-binding site of plasminogen. On the basis of its homology with HMW kininogen, the His-rich region of this protein may mediate the contact activation phase of intrinsic blood coagulation cascade.			
CC	SUBCELLULAR LOCATION: Secreted.			
CC	TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.			
CC	DOMAIN: In addition to having a high His and pro content, this protein has many internal repeats. 15 tandem repetitions of a 5-residue sequence (GH/P)[H/P]PH, (consensus) form a His/Pro-rich region.			
CC	SIMILARITY: Contains 2 cystatin-like domains.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; U32189; AAC48516.1; -			
DR	InterPro; IPR000010; Cystatin.			
DR	Pfam; PF00031; Cystatin; 1.			
DR	SMART; SM00043; Cy; 2.			
DR	Glycoprotein; Heparin-binding; Repeat; Signal.			
FT	NON_TER	1	8	POTENTIAL.
FT	CHAIN	<1	526	HISTIDINE-RICH GLYCOPROTEIN.
FT	DOMAIN	9	126	CYSTATIN-LIKE 1.
FT	DOMAIN	127	243	CYSTATIN-LIKE 2.
FT	DOMAIN	251	296	PRO-RICH.

34 182.5 26.2 485 1 SSGP VOLCA.  
35 180 25.8 1638 1 BRM DROME  
36 179 25.7 1790 1 SEPÄ EMENI  
37 176.5 25.3 1097 1 S24C ARATH  
38 175 25.1 306 1 EXTN DAUCA  
39 175 25.1 639 1 SF01 HUMAN  
40 175 25.1 653 1 SF01 MOUSE  
41 174 25.0 346 1 PRF1 LYCES  
42 173 24.8 309 1 HXA4 CHICK  
43 173 24.8 462 1 ANX7 DICDI  
44 172.5 24.7 431 1 ACRO RABIT  
45 172 24.7 134 1 PRL5 HUMAN

P21997 volvox cart  
P25439 drosophila  
P78621 emericeila  
Q9m291 arabidopsis  
P06599 caucis caro  
Q15637 homo sapien  
Q04213 mus musculu  
Q00451 lycopersico  
P17277 gallus gall  
P24639 dictyosteli  
P48038 oryctolagus  
Q99954 homo sapien





```

Db          302 HR 303

RESULT 4
PRP3 RAT
ID -PRP3 RAT      STANDARD;      PRT;      206 AA.
AC   P04474;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Acidic proline-rich protein PRP33 precursor.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RS   SEQUENCE FROM N.A.
RC   TISSUE=Parotid gland;
RX   MEDLINE=84289443; PubMed=6547951;
RA   Ziemer M.A., Swain W.F., Rutter W.J., Clements S., Ann D.K.,
RA   Carlson D.M.;
RT   "Nucleotide sequence analysis of a proline-rich protein cDNA and
RT   peptide homologues of rat and human proline-rich proteins.";
RJ   J. Biol. Chem. 259:10475-10480(1984).
RN   [2]
RS   SEQUENCE OF 1-23 FROM N.A.
RC   TISSUE=Parotid gland;
EX   MEDLINE=86033799; PubMed=3840480;
RA   Clements S., Mehanho H., Carlson D.M.;
RT   "Novel multigene families encoding highly repetitive peptide
RT   sequences. Sequence analyses of rat and mouse proline-rich protein
RT   cDNAs.";
RJ   J. Biol. Chem. 260:13471-13477(1985).
CC   -/- FUNCTION: May protect teeth by binding to tannins.
-----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC   at the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to licensese@isb-sib.ch).
-----
DR   EMBL; K02247; AAA41949.1; -.
DR   EMBL; M11898; AAA41958.1; -.
DR   PIR; A03296; FIRT3.
KW   Repeat; Parotid gland; Multigene family; Signal.
FT   SIGNAL        1      13      POTENTIAL.
FT   CHAIN         14     206     ACIDIC PROLINE-RICH PROTEIN PRP33.
FT   DOMAIN        80     189     6 X 18 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT        80     97     1.
FT   REPEAT        98    115     2.
FT   REPEAT       116    133     3.
FT   REPEAT       134    152     4.
FT   REPEAT       153    170     5.
FT   REPEAT       171    189     6.
SQ   SEQUENCE      206 AA; 21403 MW; D037592CED05CE6B CRC64;
Query Match      35.5%; Score 247.5; DB 1; Length 206;
Best Local Similarity 48.2%; Pred.No. 5.5e-11;
Matches 53; Conservative 4; Mismatches 30; Indels 23; Gaps 7;

Qy      7  RPDPHGHHPHGPPRHGHHPHGGPPPHGHP-----PHGPPRRHRPHGPPHCHPHGPP 58
      ||||| : ||||| ||||| ||||| ||||| : ||||| : |||||
Dd      79  RPPPHGNGHQRPFP-GHH-HGPSPSGPTSTSGPNPGPQQGPPQGPN-PQGPP 135
      ||||| : ||||| ||||| ||||| ||||| : ||||| : |||||
Qy      59  PHGHP-----PHGPPFRHGHHPHGGPPPHGHP--PHGFGFHDHGFCDF 97
      ||||| : ||||| ||||| ||||| ||||| : ||||| : |||||
Dd      136 PQGGPQRPPQPKPGQPPPG-QPGQPQGNPQGNPQPPQGGPQHQRPPQ 184
      ||||| : ||||| ||||| ||||| ||||| : ||||| : |||||

RESULT 5
PRP2 MOUSE

```

ID PRP2\_MOUSE STANDARD; PRT; 261 AA.  
 AC P05142;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Proline-rich protein MP-2 precursor.  
 GN PRP1 OR PRP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86059475; PubMed=2999141;  
 RA Ann D.K., Carlson D.M.;  
 RT "The structure and organization of a proline-rich protein gene of a  
 mouse multigene family.";  
 RL J. Biol. Chem. 260:15863-15872(1985).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M12099; AAA40004.1; -.  
 DR MGD; MGI:97773; Prhl.  
 KW Repeat; Signal.  
 FT SIGNAL 1 15 POTENTIAL.  
 FT CHAIN 16 261 PROLINE-RICH PROTEIN MP-2.  
 SQ SEQUENCE 261 AA; 26034 MW; 36E13BA7387F47D4 CRC64;  
 Query Match 35.4%; Score 246.5; DB 1; Length 261;  
 Best Local Similarity 43.6%; Pred. No. 7.7e-11;  
 Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;  
 YQ 7 RPPPHGHHP-----HGPPPHG-----HPPHGGPPH-----RHPPH 45  
 DB 33 RPPPSGQPRPPVNGSQGPPPPGQPRPPGPPPPGPPPPGPPPPGPPPPGPPPP 92  
 YQ 46 GPPPHG-----HPPHGGPPH-----HPPHGGPPH-----HPPHGG 86  
 DB 93 GPPPPGQPRPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPP 152  
 YQ 87 -HGFHDHGPCDPP 98  
 DB 153 PAGPQPRPPGPP 165  
 RESULT 6  
 ID\_PMP3\_MOUSE STANDARD; PRT; 296 AA.  
 AC P05143;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Proline-rich protein MP-3 (Fragment).  
 GN PRP1 OR PRP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86059475; PubMed=2999141;  
 RA Ann D.K., Carlson D.M.;  
 RT "The structure and organization of a proline-rich protein gene of a  
 mouse multigene family.";  
 RL J. Biol. Chem. 260:15863-15872(1985).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M12100; AAA40005.1; -.  
 DR MGD; MGI:97773; Prhl.  
 KW Repeat.  
 FT NON TER 1 1  
 SQ SEQUENCE 296 AA; 29521 MW; 7F146824E8AF3269 CRC64;  
 Query Match 35.4%; Score 246.5; DB 1; Length 296;  
 Best Local Similarity 43.6%; Pred. No. 8.5e-11;  
 Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;  
 YQ 7 RPPPHGHHP-----HGPPPHG-----HPPHGGPPH-----RHPPH 45  
 DB 12 RPPPSGQPRPPVNGSQGPPPPGQPRPPGPPPPGPPPPGPPPPGPPPPGPPPP 71  
 YQ 46 GPPPHG-----HPPHGGPPH-----HPPHGGPPH-----HPPHGG 86  
 DB 72 GPPPPGQPRPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPPGPPPP 131  
 YQ 87 -HGFHDHGPCDPP 98  
 DB 132 PGGPQPRPPGPP 144  
 RESULT 7  
 ID\_N075\_SOYBN STANDARD; PRT; 309 AA.  
 AC P08297;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Early nodulin 75 precursor (N-75) (NGM-75).  
 GN ENOD2A AND ENOD2B.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Wayne;  
 EX MEDLINE=91322483; PubMed=2101308;  
 RA Franssen H.J., Thompson D.V., Idler K., Kormelink R., van Kammen A.,  
 RA Bisseling T.;  
 RT "Nucleotide sequence of two soybean ENOD2 early nodulin genes  
 RT encoding Ngm-75.";  
 RL Plant Mol. Biol. 14:103-106(1990).  
 RN [2]  
 RP SEQUENCE OF 69-309 FROM N.A.  
 RA Franssen H.J., Nap J.-P., Gloudemans T., Stiekema W., van Dam H.,  
 RA Govers F., Louwerse J., van Kammen A., Bisseling T.;  
 RT "Characterization of cDNA for nodulin-75 of soybean: a gene product  
 RT involved in early stages of root nodule development.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4495-4499(1987).  
 CC -1- FUNCTION: NODULIN N-75 IS INVOLVED IN EARLY STAGES OF ROOT NODULE  
 CC DEVELOPMENT.  
 CC -1- INDUCTION: During nodulation in legume roots after Rhizobium  
 CC infection.  
 CC -1- SIMILARITY: TO OTHER PLANTS N-75.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; K03206; AAA60187.1; -.  
DR PIR; B40750; PIHUB6.  
DR Genew; HGNC:9337; PRB1.  
DR MIM; 180989; -.  
KW Repeat; Parotid gland; Multigene family; Signal.  
FT SIGNAL 1 16  
FT CHAIN 214 331 PEPTIDE IB-6.  
FT CHAIN 276 331 PEPTIDE P-H.  
FT VARIANT 106 238 Missing (in clone CP-4).  
FT VARIANT 106 258 /FTID=VAR\_005561.  
FT VARIANT 106 258 Missing (in clone CP-5).  
FT VARIANT 106 258 /FTID=VAR\_005562.  
FT CONFLICT 276 276 A -> S (IN REF. 2 AND 3).  
FT CONFLICT 331 AA; 32596 MW; 3F481FF8EBA39751 CRC64;  
SQ SEQUENCE 331 AA; 32596 MW; 3F481FF8EBA39751 CRC64;  
  
Query Match 33.5%; Score 233.5; DB 1; Length 331;  
Best Local Similarity 42.2%; Pred. No. 7e-10;  
Matches 49; Conservative 4; Mismatches 38; Indels 25; Gaps 5  
  
QY 8 PPHGHHPHPPHGHHPHGHPPH-CHPPGPPPPHPPHGHPPHGH-----PHGG 56  
Db 101 PPQGGNQPPGPPPPGKPPQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 160  
QY 57 PPHH-CHPPHGGPP-----HGHPPHGGPPHGHPPHGHPPHGHPPHGHPPH 98  
Db 161 PPQGGNQPPGPPPPGKPPQGGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 216  
  
RESULT 9  
S3B4 HUMAN  
ID S3B4 HUMAN STANDARD; PRT; 424 AA.  
AC Q15427;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Splicing factor 3B subunit 4 (Spliceosome associated protein 49) (SAP  
DE 49) (SF3B50) (Pre-mRNA splicing factor SF3B 49 kDa subunit).  
GN SF3B4 OR SAP49.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.R., Hulyk S.W.,  
RA Valladon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN CHARACTERIZATION OF THE SPLICEOSOME.  
 RP MEDLINE=20337962; PubMed=10882114;  
 RA Das R., Zhou Z., Reed R.;  
 RT "Functional association of U2 snRNP with the ATP-independent  
 RL spliceosomal complex B.";  
 RL Mol. Cell 5:779-787(2000).  
 CC -|- FUNCTION: Subunit of the splicing factor SF3B required for 'A'  
 CC complex assembly formed by the stable binding of U2 snRNP to the  
 CC branchpoint sequence (BPS) in pre-mRNA. Sequence independent  
 CC binding of SF3A/SF3B complex upstream of the branch site is  
 CC essential, it may anchor U2 snRNP to the pre-mRNA. May also be  
 CC involved in the assembly of the 'E' complex. SF3B4 has been  
 CC found in complex 'B' and 'C' as well. Belongs also to the  
 CC minor U12-dependent spliceosome, which is involved in the splicing  
 CC of rare class of nuclear pre-mRNA intron.  
 CC -|- SUBUNIT: Component of splicing factor SF3B which is composed of  
 CC four subunits; SF3B4/SAP49, SF3B3/SAP130, SF3B2/SAP145, and a  
 CC SF3B1/SAP155. SF3B associates with the splicing factor SF3A and a  
 CC 12S RNA unit to form the U2 small nuclear ribonucleoproteins  
 CC complex (U2 snRNP). SF3B4 interacts directly with SF3B2.  
 CC -|- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -|- SIMILARITY: STRONG, TO C.ELEGANS COB11.5 AND, PARTIAL, TO  
 CC YEAST HSH49.  
 CC -|- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR ENBL; L35013; AAA60300.1; -;  
 DR ENBL; BC004273; AAH04273.1; -;  
 DR ENBL; BC013886; AAH13886.1; -;  
 DR PIR; A54964; A54964.  
 DR HSP; P11940; ICDVJ.  
 DR Genew; HGNC:10771; SF3B4.  
 DR MIM; 605593; -;  
 DR GO; GO:0005681; C:spliceosome complex; TAS.  
 DR GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.  
 DR GO; GO:0006397; P:mRNA processing; TAS.  
 DR GO; GO:0006371; P:mRNA splicing; TAS.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR SMART; SM00360; RRM; 2.  
 DR PROSITE; PS01012; RRM; 2.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 KW Spliceosome; mRNA processing; mRNA splicing; Nuclear protein;  
 KW RNA-binding; Repeat.  
 FT DOMAIN 13 91 RNA-BINDING (RRM) 1.  
 FT DOMAIN 100 179 RNA-BINDING (RRM) 2.  
 FT DOMAIN 215 218 POLY-PRO.  
 FT DOMAIN 262 268 POLY-PRO.  
 SQ SEQUENCE 424 AA; 44385 MW; 212472A25D3FF002 CRC64;  
 Query Match 33.5%; Score 233.5; DB 1; Length 424;  
 Best Local Similarity 46.0%; Pred. No. 8.5e-10;  
 Matches 57; Conservative 3; Mismatches 25; Indels 39; Gaps 11;  
 QY 9 PPHGH-HPHG-PP-----PHG-HHPHGPP--HGHPGPPPPPPPHG 46  
 DB 286 PGGHSHPPPPPGMPPHPCMSQMLAHGPHGLGHFAGPPGSGGPPPPPGMPPHG 345  
 QY 47 PPHGHHPHGP---PHGHFPHG-PPHG--HPPHGPPPHGH-----PPHGHGFHDHP 94  
 DB 346 PPGMGPPRPPPGSGPMGHP--GPMPPHGMGRGPPPLMPPHGYTGRPPPPYG---YQRP 400  
 QY 95 CDDP 98  
 ||

Db 401 LPPP 404  
 RESULT 10  
 PRP2\_HUMAN  
 ID PRP2\_HUMAN STANDARD; PRT; 251 AA.  
 AC P02812;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary proline-rich protein precursor (Clone CP7) [Contains: Basic  
 DE Peptide P-F] (Fragment).  
 GN PRP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=8528325; PubMed=2993301;  
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
 RT "Differential RNA splicing and post-translational cleavages in the  
 RT human salivary proline-rich protein gene system.";  
 RL J. Biol. Chem. 260:11123-11130(1985).  
 RN [2]  
 RP SEQUENCE OF 134-194.  
 RX MEDLINE=83265674; PubMed=6874669;  
 RA Saitoh E., Isemura S., Sanada K.;  
 RT "Complete amino acid sequence of a basic proline-rich peptide, P-F,  
 RT from human parotid saliva.";  
 RL J. Biochem. 93:883-888(1983).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR ENBL; K03208; AAA60189.1; -;  
 DR PIR; S25372; PIHUPF.  
 DR Genew; HGNC:9338; PRB2.  
 DR MIM; 168810; -;  
 KW Repeat; Parotid gland; Multigene family.  
 FT NON\_TER 1 1  
 FT CHAIN 134 194 BASIC PEPTIDE P-F.  
 FT REPEAT <1 9  
 FT REPEAT 10 71  
 FT REPEAT 72 133  
 FT REPEAT 134 195  
 SQ SEQUENCE 251 AA; 24641 MW; D779F590C0BFF30B CRC64;  
 Query Match 32.9%; Score 229.5; DB 1; Length 251;  
 Best Local Similarity 43.6%; Pred. No. 1e-09;  
 Matches 48; Conservative 3; Mismatches 34; Indels 25; Gaps 5;  
 QY 8 PPHGHHPHGP---PHGHHPHGP---GHPPHGP---GHPPHGP---PPH 55  
 DB 19 PPGGNQPGPPPPPPGKPGQPPPGGNKPGQPPPGKPGQPPPGDNKSGSARSPPGK 78  
 QY 56 GPPPH-GHPHGP---PHGHHPHGP---PPHGH-----PPHGHGFHDHP 98  
 DB 79 GPPPGGNQPGPPPPPPGKPGQPPPGDNKSGSARSPPGK-----PPGPP 122  
 RESULT 11  
 PRB4\_HUMAN  
 ID PRB4\_HUMAN STANDARD; PRT; 247 AA.  
 AC P10163; P02813;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Salivary proline-rich protein PO precursor (Allele S).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the  
human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121440; PubMed=2851479;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Length polymorphisms in human proline-rich protein genes generated  
by intragenic unequal crossing over.";  
RL Genetics 120:267-278(1988).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; K03207; AAA60188.1; -;  
DR EMBL; X07882; CAA30729.1; -;  
DR FIR; S03176; PIHUSD.  
DR Genew; HGNC:9340; PRB4.  
DR MIM; 168730; -;  
DR MIM; 180990; -;  
DR Repeat; Parotid gland; Multigene family; Glycoprotein; Signal.  
FT SIGNAL 1 16  
FT CHAIN 17 247 SALIVARY PROLINE-RICH PROTEIN PO.  
FT CHAIN 17 39 PROTEIN NL.  
FT CHAIN 40 177 GLYCOSYLATED PROTEIN A.  
FT CONFLICT 37 57 MISSING (IN REF. 2).  
FT CONFLICT 218 218 D -> A (IN REF. 2).  
SQ SEQUENCE 247 AA; 25108 MW; 6A1943B435161691 CRC64;  
  
Query Match 32.7%; Score 228; DB 1; Length 247;  
Best Local Similarity 36.4%; Pred. No. 1.3e-09;  
Matches 52; Conservative 4; Mismatches 39; Indels 48; Gaps 5;  
  
QY 7 RPPPHGHHPGPPPHGHPGPPPHG-----HPPHGGPPPHGPPPHG-----PPPHGH 52  
Db 81 RPPQGNQSGPPPHGPKERPPPPQGGNQSQTTPPPGKPERPPQGGNQSHPRPPPGK 140  
QY 53 P-----PHGPPPHGHPGPPPHG-----HPPHGGPPPHGPPPHG-----HPPHGGPPPH 80  
Db 141 PERPPQGGNQSQTTPPPHGGPKERPPPPQGGNQSQTTPPPGKPERPPQGGNQSHPRPPPGK 200  
QY 81 GHP--PHGHPGHPGPPPHGPPPHG 101  
Db 201 GKQGGPPAGNQQQDPPAGK 223  
  
RESULT 12  
PRPL HUMAN  
ID PRPL HUMAN STANDARD; PRT; 276 AA.  
AC P10162; P02813;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO (Allele K) [Contains: Peptide P-D]  
DE (Fragment).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121439; PubMed=3220251;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Many protein products from a few loci: assignment of human salivary  
RT proline-rich proteins to specific loci.";  
RL Genetics 120:255-265(1988).  
RN [2]  
RP SEQUENCE OF 207-276  
RX MEDLINE=83186122; PubMed=6841349;  
RA Saitoh E., Isemura S., Sanada K.;  
RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
RT from human parotid saliva.";  
RL J. Biochem. 93:495-502(1983).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X07715; CAA30543.1; ALT\_SEQ.  
DR FIR; S03176; PIHUSD.  
DR MIM; 168730; -;  
DR MIM; 180990; -;  
DR GO; GO:0005576; C:extracellular; NAS.  
DR InterPro; IPR000637; AT\_hoek.  
DR PRINTS; PR00929; ATHOOK.  
KW Repeat; Parotid gland; Multigene family.  
FT REPEAT 1  
FT CHAIN 207 276 PEPTIDE P-D.  
SQ SEQUENCE 276 AA; 27816 MW; 9F49426C979441A CRC64;  
  
Query Match 32.1%; Score 224; DB 1; Length 276;  
Best Local Similarity 37.6%; Pred. No. 2.7e-09;  
Matches 50; Conservative 3; Mismatches 36; Indels 44; Gaps 5;  
  
QY 7 RPPPHGHHPGPPPHGHPGPPPHG-----HPPHGGPPPHGPPPHG-----HPPHGGHP 53  
Db 131 RPPQGNQSGPPPHGPKERPPPPQGGNQSQTTPPPGKPERPPQGGNQSHPRPPPGK 190  
QY 54 PHGPP-----HPPHGGPPPHGHPGPPPHG-----HPPHGGPPPHG-----PPH 85  
Db 191 PEGPPQGGNKSARSPPGKPPQGGNQSQTTPPPGKPERPPQGGNQSHPRPPPGK 250  
QY 86 GHGFHDHGGCDPP 98  
Db 251 G---KPGGPPPP 260  
  
RESULT 13  
PRPM HUMAN  
ID PRPM HUMAN STANDARD; PRT; 234 AA.  
AC P10161; P02813;  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Salivary proline-rich protein PO (Allele M) [Contains: Peptide P-D]  
DE (Fragment).  
GN PRB4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89121439; PubMed=3220251;  
RA Lyons K.M., Stein J.H., Smithies O.;  
RT "Many protein products from a few loci: assignment of human salivary

RT proline-rich proteins to specific loci.;  
 RL Genetics 120:255-265(1988).  
 RN [2]  
 RP SEQUENCE OF 165-234.  
 RX MEDLINE=83186122; PubMed=6841349;  
 RA Saitoh E., Isemura S., Sanada K.;  
 RT "Complete amino acid sequence of a basic proline-rich peptide, P-D,  
 RL from human parotid saliva.";  
 RT J. Biochem. 93:495-502(1983).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X07704; CAA30542.1; -.  
 DR MIM; 168730; -.  
 DR MIM; 180930; -.  
 DR GO; 0005576; C:extracellular; NAS.  
 KW Repeat; Parotid gland; Multigene family.  
 FT NON\_TER 1  
 FT CHAIN 165 234  
 FT PEPTIDE P-D.  
 FT SEQUENCE 234 AA; 23676 MW; 310AFF13A44E747F CRC64;  
 SQ  
 Query Match 31.8%; Score 221.5; DB 1; Length 234;  
 Best Local Similarity 42.3%; Pred. No. 3.4e-09;  
 Matches 47; Conservative 5; Mismatches 40; Indels 19; Gaps 5;  
 QY 7 RPPPHGHHGPPPH-GHPHGGPPPHGPPHGGPPH-HPPHGGPPH-----GH 52  
 DB 15 RPPPPKQPPPPGQGNQSQGPPPPGKPPPPGQGNQSQGPPPPGKPPPPGQGN 74  
 QY 53 PPHGPPPHGPPHGGPPH-GHPHGGPPH-----GHPHGGHFDGCDPP 98  
 DB 75 QSQGPPPTGKPPGPPPPGQGNQSQGPPPPGKPPPPGQGNQSHRPPPP 125  
 RESULT 14  
 ID NO75 LUPLU STANDARD; PRT; 434 AA.  
 AC Q0681; 1  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Early nodulin 75 protein (N-75) (NGM-75) (Fragment).  
 GN ENOD2.  
 OS Lupinus luteus (Yellow lupine).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.  
 CC NCBI\_TaxID=3873;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root nodules;  
 RX MEDLINE=91355883; PubMed=2103455;  
 RA Szczyslowski K., Legocki A.B.;  
 RT "Isolation and nucleotide sequence of cDNA clone encoding nodule-  
 RT specific (hydroxy)proline-rich protein LENOD2 from yellow lupin.";  
 RL Plant Mol. Biol. 15:361-363(1990).  
 CC -!- FUNCTION: INVOLVED IN EARLY STAGES OF ROOT NODULE DEVELOPMENT.  
 CC -!- SIMILARITY: TO OTHER PLANTS N-75.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; X55371; CAA39050.1; -.  
 DR PIR; S11967; S11967.  
 KW Nodulation; Repeat.  
 FT NON\_TER 1  
 FT DOMAIN 6 407  
 FT SEQUENCE 434 AA; 49827 MW; 8CE8E140C21B6664 CRC64;  
 SQ  
 Query Match 30.2%; Score 210.5; DB 1; Length 434;  
 Best Local Similarity 37.0%; Pred. No. 3.1e-08;  
 Matches 51; Conservative 6; Mismatches 40; Indels 41; Gaps 8;  
 QY 4 IIRPPPHG-----HHPHG-----PPPHGHP-HGPPPHGHPHGGPPPH-----P 43  
 DB 137 LVH-PPPHDPPPIYHPPHKKPIYPPHKKPIYPPHKKPIYPPHKKPIYPPHKKPVYEP 195  
 QY 44 PHGPPPHGHP-----HGPPPHGHPHGGPPPHGHPHGGPPPHGHPHGGPPPH 84  
 DB 196 PYEKPPVPPHPPDEKPPPIYPPHKKPIYPPHKKPIYPPHKKPIYPPHKKPIYPPHKKPP 255  
 QY 85 -HGHPHGGHGGCDPPSHK 101  
 DB 256 VYEPYERSPPVHPPSHE 273  
 RESULT 15  
 ID EXTN TOBAC STANDARD; PRT; 620 AA.  
 AC P13983;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Extensin precursor (Cell wall hydroxyproline-rich glycoprotein).  
 GN HRGPNT3.  
 OS Nicotiana tabacum (Common tobacco).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 CC lamids; Solanales; Solanaceae; Nicotiana.  
 CC NCBI\_TaxID=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi; TISSUE=Leaf;  
 RX MEDLINE=90128263; PubMed=2612909;  
 RA Keller B., Lamb C.J.;  
 RT "Specific expression of a novel cell wall hydroxyproline-rich  
 RT glycoprotein gene in lateral root initiation.";  
 RL Genes Dev. 3:1639-1646(1989).  
 CC -!- FUNCTION: Has a specialized structural function, possibly in  
 CC the mechanical penetration of the cortex and epidermis of the  
 CC main root.  
 CC -!- SUBCELLULAR LOCATION: Extracellular matrix.  
 CC -!- FTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
 CC GLYCOSYLATED.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X13885; CAA32090.1; -.  
 DR PIR; S06733; S06733.  
 DR InterPro; IPR000480; Glutelin.  
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 FT HYDROXYLATION.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 620 EXTENSIN.  
 FT REPEAT 70 73 H-A-P-P.  
 FT REPEAT 148 151 H-A-P-P.  
 FT

FT DOMAIN 229 242 2 X 7 AA TANDEM REPEATS OF T-H-R-H-A-P-P.  
FT REPEAT 229 235 1.  
FT REPEAT 236 242 2.  
FT DOMAIN 205 620 CONTAINS THE SER-PRO(4) REPEATS.  
FT DOMAIN 499 600 3 X APPROXIMATE TANDEM REPEATS.  
SQ SEQUENCE 620 AA; 65406 MW; 641DD2278AB28524 CRC64;  
Query Match 30.1%; Score 210; DB 1; Length 620;  
Best Local Similarity 42.3%; Pred. No. 4.4e-08;  
Matches 55; Conservative 4; Mismatches 29; Indels 42; Gaps 13;  
QY 8 PPHGH-----HPCPPPHGHP---HCPPPH-GH-PP---HGPPPR---HPPHG 46  
DB 59 PPSRGHVPSPRHAPPRHAYPPSHGHLPPSVGGPPPHRGHLPPSRGFNPPSPSVISPSHP 118  
QY 47 PPHGHPP--HGP---PPGHPP-----HCPPPHGH-PPHPPPHGH-----PPHGHGFH 90  
DB 119 PPSYGAPPSPSHGHLPPSHGQRPSPSPSHGHAPPSCGHTPPRGQHPHRRPPSPSRHG-- 176  
QY 91 DHGPCDPPSH 100  
DB 177 -HPP--PTY 183

Search completed: September 23, 2004, 22:50:36  
Job time : 12.7631 secs

Blank



Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	420.5	60.3		546	6	Q9BGU1	Q9bgl bos taurus
2	375	53.8		525	11	Q9PS8U	Q9ps8 rattus norv
3	333	47.8		525	11	Q9PS6U	Q9ps6 mus musculu
4	333	47.8		525	11	Q9EB3U	Q9eb3 mus musculu
5	333	47.8		525	11	Q9PS5U	Q9ps5 mus musculu
6	323	46.3		515	11	Q9PS7U	Q9ps7 rattus norv
7	315	45.2		510	11	Q9EB2U	Q9eb2 rattus norv
8	259.5	37.2		295	11	Q07611	Q07611 rattus norv
9	258.5	37.1		511	6	Q95JD0	Q95jd0 sus scrofa
10	258.5	37.1		566	6	Q95JD1	Q95jd1 sus scrofa
11	254.5	36.5		676	6	Q95JC9	Q95jc9 sus scrofa
12	246.5	35.4		227	11	Q62107	Q62107 mus musculu
13	246.5	35.4		317	11	Q62103	Q62103 mus musculu
14	246	35.3		274	11	Q04154	Q04154 rattus norv
15	242	34.7		112	11	Q9CR74	Q9cr74 mus musculu
16	236	33.9		330	10	Q41402	Q41402 sesbania ro

397 HHPHGHHPHGHHPYGHHPHGHHPHNDYDHPGCDPPHR 441

4:

OS *Rattus norvegicus* (Rat).

QY	54	PHGPPPHGHPGPPHGP-PPHGHPPHGHPPHGHPPHGH 86	
Db	343	PPGPPPPGPPPGAPPGAPPGAPPPGPPPPGPPPPG 376	
RESULT 11			
Q95JC9		PRELIMINARY; PRT; 676 AA.	
ID	Q95JC9		
AC	Q95JC9		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Basic proline-rich protein.		
OS	Sus scrofa (Pig).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
OX	NCBI_TaxID=9823;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Parotid gland;		
RA	Zhang Q., Szalay A.A., Kyeune-Nyombi E., Sands J.F., Oberg K.C.,		
RA	Tieche J.-M., Leonora J.;		
RT	"Cloning and expression of a novel proline-rich protein from porcine		
RT	parotid glands.";		
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
RL	EMBL; AY035849; AAK61383.1; -		
DR	GO; GO:0016021; C:integral to membrane; IEA.		
DR	GO; GO:0005249; P:voltage-gated potassium channel activity; IEA.		
DR	GO; GO:0006813; P:potassium ion transport; IEA.		
DR	InterPro; IPR005404; KV3.3 channel.		
DR	PRINTS; PRO1582; KV33CHANNEL.		
DR	SEQUENCE 676 AA; 62297 MW; 3008BC41EFD81FC9 CRC64;		
Query Match 36.5%; Score 254.5; DB 6; Length 676;			
Best Local Similarity 50.9%; Pred. No. 9.5e-17; Indels 15; Gaps 4;			
Matches 54; Conservative 0; Mismatches 37;			
QY	8	PPPHGHHPHGP-PPHGHHPHGHPPHGHHPHGH-----PPRHPHGHPPHGH-----HP 53	
Db	314	PPPPPPPPGAPPGAPPPGPPPPGPPPPGAPPGAPPPGPPPPGPPPPGAPPGAPPP 419	
QY	54	PHGPPPHGHPGHP-PPHGHPPHGHPPHGHPPHGHGHFHGHGCDPP 98	
Db	374	PPGPPPPGPPPGAPPGAPPPGPPPPGPPPPGPPPGAPPGAPPGAPPGAPPPPP 419	
RESULT 12			
Q62107		PRELIMINARY; PRT; 227 AA.	
ID	Q62107		
AC	Q62107		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Proline-rich salivary protein (Fragment).		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=8603799; PubMed=3840480;		
RA	Clements S., Mehansho H., Carlson D.M.;		
RT	"Novel multigene families encoding highly repetitive peptide		
RT	sequences: Sequence analyses of rat and mouse proline-rich protein		
RT	CDNAs.";		
RL	J. Biol. Chem. 260:13471-13477(1985).		
RL	EMBL; M11902; AAA40003.1; -		
DR	PIR; C29149; C29149.		
DR	NON_TER 227		
DR	SEQUENCE 227 AA; 22822 MW; D893B884C76BA7BD CRC64;		
QY		35.4%; Score 246.5; DB 11; Length 227;	
Best Local Similarity 35.4%; Pred. No. 2.1e-16;			

Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;

QY 7 RPPPHGHP-----HPPPHG-----HPPHPPPHG-----HPPH 45  
DB 33 RPPPSGQPRPVNGSQGPPPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 92  
QY 46 GPPPHG-----HPPHPPPHG-----HPPHPPPHG-----HPPH 86  
DB 93 GPPPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 152

QY 87 -HGFDHGPCDPP 98  
DB 153 PAGQPRPPQGP 165

RESULT 13

Q62103 PRELIMINARY; PRT; 317 AA.

AC Q62103;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Proline-rich protein precursor.  
GN PRP2 OR PRP.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]

SEQUENCE FROM N.A.  
RC STRAIN=CD-1; TISSUE=Liver;  
RX MEDLINE=88273214; PubMed=2839509;  
RA Ann D.K., Smith M.K., Carlson D.M.;  
RT "Molecular evolution of the mouse proline-rich protein multigene family. Insertion of a long interspersed repeated DNA element.";  
RL J. Biol. Chem. 263:10887-10893(1988).  
DR EMBL; M23236; AAA53048.1; -  
DR PIR; A28996; A28996.  
DR MGD; MGI:1932491; Prp2.  
KW Signal.  
FT SIGNAL.  
FT CHAIN  
SQ SEQUENCE 317 AA; 31719 MW; 019301BE31D73278 CRC64;

Query Match 35.4%; Score 246.5; DB 11; Length 317;  
Best Local Similarity 43.6%; Pred. No. 2.8e-16;  
Matches 58; Conservative 0; Mismatches 34; Indels 41; Gaps 9;

QY 7 RPPPHGHP-----HPPPHG-----HPPHPPPHG-----HPPH 45  
DB 33 RPPPSGQPRPVNGSQGPPPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 92  
QY 46 GPPPHG-----HPPHPPPHG-----HPPHPPPHG-----HPPH 86  
DB 93 GPPPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 152

QY 87 -HGFDHGPCDPP 98  
DB 153 PAGQPRPPQGP 165

RESULT 14

Q04154 PRELIMINARY; PRT; 274 AA.

AC Q04154;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Salivary proline-rich protein.  
GN RP15.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

NCBI\_TaxID=101116;  
[1]

SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=91257817; PubMed=2045095;  
RA Lin H.H., Ann D.K.;  
RT "Molecular characterization of rat multigene family encoding proline-rich proteins.";  
RL Genomics 10:102-113(1991).  
DR EMBL; M64793; AAA42064.1; -  
DR PIR; B39066; B39066.  
SQ SEQUENCE 274 AA; 27380 MW; 6A57121F8F07387B CRC64;

Query Match 35.3%; Score 246; DB 11; Length 274;  
Best Local Similarity 47.1%; Pred. No. 2.8e-16;  
Matches 57; Conservative 1; Mismatches 29; Indels 34; Gaps 9;

QY 8 PPHG--HPPHPPPHG--HPPHPPPHG--HPPHPPPHG--HPPHPPPHG--HPPH 53  
DB 104 PPGGQPRPPGPPPGGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 163  
QY 54 --PHGPPPHG--HPPHPPPHG--HPPHPPPHG--HPPHPPPHG--HPPHPPPHG-- 97  
DB 164 QGPGPPPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPPGQPRPP 219

QY 98 P 98  
DB 220 P 220

RESULT 15

Q9CR74 PRELIMINARY; PRT; 112 AA.

AC Q9CR74;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE EST A238765.  
GN 1810007E14RIK OR A238765.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]

SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Pancreas, and Small intestine;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shiragawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL; AK019007; BAB31507.1; -  
DR EMBL; AK008311; BAB25594.1; -  
DR MGD; MGI:1858382; 1810007E14RIK.  
SQ SEQUENCE 112 AA; 10637 MW; 16D773538A95A230 CRC64;

Query Match 34.7%; Score 242; DB 11; Length 112;

Best Local Similarity 50.5%; Pred. No. 3e-16;  
Matches 54; Conservative 0; Mismatches 39; Indels 14; Gaps 6;

QY 6 HRPPTGHHPHGPPPHGHHPHGP-----PPHGHPHPHPPPPHHP-HGPPPHGH-PPHGPPP 59  
Db 8 HCGPPDPGHP-CHCEGHHPFPFGGPHPGH-PGCHPGGHHPPCHGSPSHGCHPGCHPGG 65

QY 60 HGHHPGHPHPGHP-----HGPPPHGHPHGHGPHDHGPPCDPPSH 100  
Db 66 HHPPHGGSGHGHGPHGCPGGHHPPGHGGPSSHGGSGHCHGCHPFGPH 112

Search completed: September 23, 2004, 22:55:05  
Job time : 61.6265 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 23, 2004, 22:47:25 ; Search time 18,253 Seconds  
(without alignments)  
285.864 Million cell updates/sec

Title: US-10-074-225A-6

Perfect score: 697

Sequence: 1 SVNTIHRPPHGHHPHGHPPP.....RPPHGHGFHDHGPCDPPSHK 101

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pdp.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pdp.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pdp.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pdp.\*
- 5: /cgn2\_6/ptodata/2/iaa/6C COMB.pdp.\*
- 6: /cgn2\_6/ptodata/2/iaa/6D COMB.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	314	45.1	525	4	US-09-976-594-64
2	229.5	32.9	124	2	US-08-925-237-2
3	215.5	30.9	104	4	US-08-547-693-235
4	208	29.8	865	4	US-08-281-766-19
5	196	28.1	82	2	US-08-925-237-4
6	175	25.1	306	1	US-08-217-327-6
7	171	24.5	214	1	US-08-217-327-4
8	167	24.0	1958	1	US-07-945-283-2
9	164	23.5	167	3	US-09-507-323B-13
10	161	23.1	79	1	US-08-487-359-3
11	161	23.1	79	1	US-08-222-798A-3
12	160.5	23.0	78	1	US-08-487-359-5
13	160.5	23.0	78	1	US-08-222-798A-5
14	160.5	23.0	144	1	US-08-642-255-49
15	160.5	23.0	720	3	US-09-219-849-4
16	160.5	23.0	777	1	US-08-642-255-53
17	160	23.0	79	1	US-08-487-359-4
18	160	23.0	79	1	US-08-222-798A-4
19	159	22.8	79	1	US-08-487-359-1
20	159	22.8	79	1	US-08-487-359-8
21	159	22.8	79	1	US-08-222-798A-1
22	159	22.8	79	1	US-08-222-798A-8
23	158.5	22.7	148	4	US-09-461-325-453
24	158.5	22.7	148	4	US-10-012-542-453
25	158.5	22.7	234	1	US-08-642-255-51
26	158.5	22.7	504	3	US-08-642-255-51
27	158.5	22.7	561	1	US-08-642-255-52

28	158	22.7	79	1	US-08-487-359-2	Sequence 2, Appli
29	158	22.7	79	1	US-08-487-359-6	Sequence 6, Appli
30	158	22.7	79	1	US-08-222-798A-2	Sequence 2, Appli
31	158	22.7	79	1	US-08-222-798A-6	Sequence 6, Appli
32	157	22.5	76	4	US-09-547-693-233	Sequence 233, App
33	155	22.2	590	1	US-08-021-608D-8	Sequence 8, Appli
34	155	22.2	590	1	US-08-726-160-8	Sequence 8, Appli
35	155	22.2	590	5	PCT-US94-01782-8	Sequence 8, Appli
36	155	22.2	643	1	US-08-021-608D-10	Sequence 10, Appl
37	155	22.2	643	5	US-08-726-160-10	Sequence 10, Appl
38	155	22.2	643	5	PCT-US94-01782-10	Sequence 10, Appl
39	155	22.2	644	1	US-08-021-608D-2	Sequence 2, Appli
40	155	22.2	644	1	US-08-726-160-2	Sequence 2, Appli
41	155	22.2	644	5	PCT-US94-01782-2	Sequence 2, Appli
42	151.5	21.7	330	1	US-08-642-255-32	Sequence 32, Appl
43	151.5	21.7	408	1	US-07-609-716-65	Sequence 65, Appl
44	151.5	21.7	408	3	US-08-475-411A-65	Sequence 65, Appl
45	151.5	21.7	408	3	US-08-478-029A-65	Sequence 65, Appl

## ALIGNMENTS

### RESULT 1

US-09-976-594-64  
; Sequence 64, Application US/09976594  
; Patent No. 6673549  
; GENERAL INFORMATION:  
; APPLICANT: Furness, Michael  
; APPLICANT: Buchbinder, Jenny  
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS  
; FILE REFERENCE: PA-0041 US  
; CURRENT APPLICATION NUMBER: US/09/976,594  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: 60/240,409  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 1143  
; SOFTWARE: PERL Program  
; SEQ ID NO 64  
; LENGTH: 525  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6673549 085596CD1  
US-09-976-594-64

Query Match 45.1%; Score 314; DB 4; Length 525;  
Best Local Similarity 41.8%; Pred. No. 3.7e-21;  
Matches 64; Conservative 2; Mismatches 23; Indels 64; Gaps 8;

QY 7 RPP-PHG---HHPGPPPHGHPHHPGPP---HGHP-PHGPPP-----40

Db 275 KPPEKPGSRDHHHPKP-----HEGPPPPPPDERDHSHPPLPQGPPLPMSCCSCQH 329

QY 41 -----RHP-----PHGPPPHGHP-----HGPPPHGHP 67

Db 330 ATFTGCAQRSSNNSSDLPHPKHSHEQHGHHPHHPHHPHHPHHPHHPHHPHHPH 389

QY 68 PPHGHPGPPPHGHPHHPGHPHHPGHPHHPGHPHHPGHPHHPGHPHHPGHPHHP 100

Db 330 HPHGHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPHHPH 422

### RESULT 2

US-08-925-237-2  
; Sequence 2, Application US/08925237  
; Patent No. 5981720  
; GENERAL INFORMATION:  
; APPLICANT: Azen, Edwin A.  
; APPLICANT: Pan, David  
; TITLE OF INVENTION: Human Salivary Proteins And Fragments  
; THEREOF HAVING ALPHA-GLUCOSIDASE INHIBITORY ACTIVITY









QY	67	PPPHGHPHGP	83
DB	63	PPP---PPFRPPFRGP	76

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

ATTORNEY/AGENT INFORMATION:  
NAME: MURASHIGE, KATE H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 2000-0553.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 887-1500  
TELEFAX: (202) 887-0763  
TELEX: 90-4030  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 79 amino acids

[illegible]

```

; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/POCKET NUMBER: 2000-0553.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-222-798A-5

Query Match      23.0%; Score 160.5; DB 1; Length 78;
Best Local Similarity 46.8%; Pred.No. 3.7e-08;
Matches 36; Conservative 3; Mismatches 33; Indels 5; Gaps 3;

QY      8 PPHGHHPGPPRHGHHPGGPRRHRP-PHGPPRHGHHPRHGHHPHG 66
       |||
Db      3 PPNVVGPRFPFGFRFPFPFPFPFPFPFPFPFPFPFPFPFPFP 61
       |||

QY      67 PPHGHHPGPPPHGHP 83
       |||
Db      62 PPP---PPFPFPFPFGPP 75
       |||

```

```

RESULT 14
US-08-642-255-49
; Sequence 49, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 PHT UR
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-642-255-49
Query Match 23.0%; Score 160.5; DB 1; Length 144;
Best Local Similarity 38.7%; Pred. NO. 6.5e-08;

```

```

Matches 55; Conservative 0; Mismatches 36; Indels 51; Gaps 13
QY 6 HRP--PPHGHHPHGP-----PPHGHHPHGPP-----PHGHP-PHGPP-----PRHP 43
DB 3 HGPAGPKGAHGPAGPKGAHGPAGPCAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 62
QY 44 PHGPPPHGHP-----PHGPP-----PHGHP-PHGPP-----PHGHP-PHGPP-----P 79
DB 63 PGAGPGAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 122
QY 80 HGHPPHGHG-----FHDHGCPDP 97
DB 123 AG-PPGAHGPAGPKGAHGPAGP 143

RESULT 15
US-09-219-849-4
; Sequence 4, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUTSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; TITLE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-4

Query Match 23.0%; Score 160.5; DB 3; Length 720;
Best Local Similarity 38.7%; Pred. No. 2.8e-07; Indels 51; Gaps 13;
Matches 55; Conservative 0; Mismatches 36; Indels 51; Gaps 13;
QY 6 HRP--PPHGHHPHGP-----PPHGHHPHGPP-----PHGHP-PHGPP-----PRHP 43
DB 3 HGPAGPKGAHGPAGPKGAHGPAGPCAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 62
QY 44 PHGPPPHGHP-----PHGPP-----PHGHP-PHGPP-----PHGHP-PHGPP-----P 79
DB 63 PGAGPGAGPPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 122
QY 80 HGHPPHGHG-----FHDHGCPDP 97
DB 123 AG-PPGAHGPAGPKGAHGPAGP 143

```

Search completed: September 23, 2004, 22:56:49  
Job time : 19.253 secs



[illegible]

[REDACTED]

7 RPPRHGHHRRGGRRHGHGHRPHGHPGRH-PRH-GPPRHGHPRH-GPPRHCHP 63  
95 KPPRPGQRRHKKPRH---ENPPRHQRHKEKPRHQRHKEKPRHQRHKEKPRHQRH 150





Blank